Electronic Colloquium on Computational Complexity, Report No. 40 (2021)

Majority vs. Approximate Linear Sum and Average-Case Complexity Below NC^1

Lijie Chen* Zhenjian Lu† Xin Lyu‡ Igor C. Oliveira§

March 15, 2021

Abstract

We develop a general framework that characterizes strong average-case lower bounds against circuit classes $\mathcal{C}$ contained in NC^1, such as $\mathsf{AC}^0[\oplus]$ and ACC^0. We apply this framework to show:

- *Generic seed reduction*: Pseudorandom generators (PRGs) against $\mathcal{C}$ of seed length $\leq n-1$ and error $\varepsilon(n) = n^{-\omega(1)}$ can be converted into PRGs of *sub-polynomial* seed length.
- *Hardness under natural distributions*: If E (deterministic exponential time) is average-case hard against $\mathcal{C}$ under *some* distribution, then E is average-case hard against $\mathcal{C}$ under the *uniform* distribution.
- *Equivalence between worst-case and average-case hardness*: Worst-case lower bounds against $\mathsf{MAJ} \circ \mathcal{C}$ for problems in E are *equivalent* to strong average-case lower bounds against $\mathcal{C}$. This can be seen as a certain converse to the Discriminator Lemma [Hajnal et al., JCSS'93].

These results were not known to hold for circuit classes that do not compute majority. Additionally, we prove that classical and recent approaches to *worst-case* lower bounds against ACC^0 via communication lower bounds for NOF multi-party protocols [Håstad and Goldmann, CC'91; Razborov and Wigderson, IPL'93] and Torus polynomials degree lower bounds [Bhrushundi et al., ITCS'19] also imply *strong average-case hardness* against ACC^0 under the uniform distribution.

Crucial to these results is the use of *non-black-box* hardness amplification techniques and the interplay between ***Majority*** (MAJ) and ***Approximate Linear Sum*** ($\widetilde{\mathsf{SUM}}$) gates. Roughly speaking, while a MAJ gate outputs 1 when the sum of the m input bits is at least $m/2$, a $\widetilde{\mathsf{SUM}}$ gate computes a real-valued bounded weighted sum of the input bits and outputs 1 (resp. 0) if the sum is close to 1 (resp. close to 0), with the promise that one of the two cases always holds. As part of our framework, we explore ideas introduced in [Chen and Ren, STOC'20] to show that, for the purpose of proving lower bounds, a top layer MAJ gate is *equivalent* to a (weaker) $\widetilde{\mathsf{SUM}}$ gate. Motivated by this result, we extend the algorithmic method and establish stronger lower bounds against bounded-depth circuits with layers of MAJ and $\widetilde{\mathsf{SUM}}$ gates. Among them, we prove that:

- *Lower bound*: NQP does not admit fixed quasi-polynomial size $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$ circuits.

This is the first explicit lower bound against circuits with distinct layers of MAJ, $\widetilde{\mathsf{SUM}}$, and THR gates. Consequently, if the aforementioned equivalence between MAJ and $\widetilde{\mathsf{SUM}}$ as a *top gate* can be extended to *intermediate layers*, long sought-after lower bounds against the class $\mathsf{THR} \circ \mathsf{THR}$ of depth-2 polynomial-size threshold circuits would follow.

*Massachusetts Institute of Technology, USA. lijieche@mit.edu
†University of Warwick, UK. zhen.j.lu@warwick.ac.uk
‡Tsinghua University, China. lvx17@mails.tsinghua.edu.cn
§University of Warwick, UK. igor.oliveira@warwick.ac.uk

ISSN 1433-8092

Contents

1 Introduction **3**
- 1.1 Overview . . . 3
- 1.2 Results and techniques . . . 4
  - 1.2.1 Equivalences for worst-case and strong average-case hardness . . . 4
  - 1.2.2 Lower bounds against circuits with layers of $\widetilde{\mathsf{SUM}}$ and MAJ gates . . . 9

2 Preliminaries **12**
- 2.1 Notation . . . 12
- 2.2 A $\oplus\mathsf{L}$-complete problem with good properties . . . 12

3 Equivalences for worst-case and strong average-case lower bounds **13**
- 3.1 Preliminaries . . . 13
- 3.2 Proof of Theorem 1 . . . 15
- 3.3 A weak equivalence theorem in the constant-error regime . . . 22

4 Lifting worst-case ACC^0 lower bound approaches to strong correlation bounds **23**
- 4.1 Preliminaries . . . 23
- 4.2 $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0$ as torus polynomials . . . 24
- 4.3 Proof of Theorem 2 . . . 27

5 Lower bounds against circuits with MAJ, THR, and $\widetilde{\mathsf{SUM}}$ gates **27**
- 5.1 Preliminaries . . . 28
- 5.2 $\mathsf{LTF}^{\mathrm{quasipoly}(n)} \circ \mathcal{C}$ lower bounds . . . 31
- 5.3 Average-case lower bounds against $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$ . . . 33
  - 5.3.1 Certifying low-depth circuits: Proof of Lemma 26 . . . 34

1 Introduction

1.1 Overview

Establishing the *intractability of computations* and understanding the *power of randomness* in algorithms are among the most basic open problems in theoretical computer science. The theory of computational pseudorandomness provides a firm link between these two research directions. One of the most celebrated developments in this area is a proof that if E (deterministic exponential time $2^{O(n)}$) requires Boolean circuits of exponential size then $\mathsf{P} = \mathsf{BPP}$ [IW97, STV01]. This result and its underlying techniques provide a robust mathematical theory that connects *worst-case lower bounds*, *average-case hardness*, and the construction of *pseudorandom generators*.

Unfortunately, a large part of this beautiful and far-reaching theory is not known to survive in *restricted* computational settings. For instance, while we know since the eighties that E cannot be $(1/2 + n^{-1/2+\Omega(1)})$-approximated by $\mathsf{AC}^0[\oplus]$ [Raz87], it is an important open problem to obtain *strong* average-case hardness results of the form $1/2 + n^{-k}$ for all k and pseudorandom generators against this circuit class. The fact that existing connections between hardness and pseudorandomness do not apply in restricted settings is significant, given that known unconditional results and existing lower bound frontiers lie within weak sub-classes of NC^1, such as ACC^0.

Several works (e.g. [Vio05, GR08, SV10, LTW11, AS14, GSV18, Vio19, IM21]) have investigated the difficulty of extending the hardness vs. randomness theory and its consequences to restricted circuit classes. Roughly speaking, these results show that standard "black-box" techniques to amplify computational hardness and construct pseudorandom generators *require* the underlying circuit class $\mathcal{C}$ to be closed under *majority*. However, obtaining lower bounds against circuit classes that are closed under majority is a notorious open problem. This leaves us in this unsatisfying situation where many benefits of the theory mentioned above only apply to settings where current circuit-analysis techniques do not hold. In other words, we have the following "lose-lose" scenario: above TC^0 we have no lower bounds, while below it we have lower bounds but no hardness amplification.

In this work, we explore *non-black-box* techniques to overcome this difficulty, obtaining a general connection between worst-case lower bounds, strong average-case hardness, and pseudorandomness for *weak circuit classes*. Our results build on recent ideas of Chen and Ren [CR20] employed in the context of the algorithmic method. Using our techniques, we are able to establish fundamental equivalences that were previously only known for circuit classes containing TC^0. As a consequence, the new results are widely applicable and can affect *current frontiers in circuit complexity theory*.

A crucial ingredient in our proofs is the interplay between Majority (MAJ) and Approximate Linear Sum ($\widetilde{\mathsf{SUM}}$) gates. Roughly speaking, while a MAJ gate outputs 1 when the sum of the m input bits is at least $m/2$, a $\widetilde{\mathsf{SUM}}$ gate computes a real-valued bounded weighted sum of the input bits and outputs 1 (resp. 0) if the sum is close to 1 (resp. close to 0), with the promise that one of the two cases always holds. $\widetilde{\mathsf{SUM}}$ gates are significantly simpler than MAJ gates (e.g. MAJ has approximate degree [NS94] of order $\Omega(m)$), but still powerful enough to implement useful computations, such as hardness amplification for *specific* problems (a non-black-box element).

Complementing our results about the average-case complexity of restricted circuit classes, we obtain the first unconditional lower bounds against bounded-depth circuits with distinct layers of MAJ, $\widetilde{\mathsf{SUM}}$, and THR gates. These results suggest that further investigating the relation between MAJ and $\widetilde{\mathsf{SUM}}$ might be a path to lower bounds against depth-2 threshold circuits, a long-standing open problem in complexity theory (cf. [GHR92, CM18]).

1.2 Results and techniques

To begin with, we recall some definitions for linear sums of functions. Our notation is taken from previous work [Wil18a, CW19, CR20, CLW20] on lower bounds via the algorithmic method. Let $\mathcal{C}$ be a class of functions from $\{0,1\}^n \to \{0,1\}$.

$\mathsf{SUM} \circ \mathcal{C}$-circuits. A $\mathsf{SUM} \circ \mathcal{C}$-circuit $C\colon \{0,1\}^n \to \mathbb{R}$ is a circuit that can be written as $C(x) = \sum_{i=1}^{\ell} \alpha_i \cdot C_i(x)$, where each α_i is a real, and each $C_i \in \mathcal{C}$. Here ℓ is called the *sparsity* of C, and is denoted as $\mathsf{sparsity}(C)$. We also use $\mathsf{complexity}(C)$ to denote $\max(\ell, \sum_{i=1}^{\ell} |\alpha_i|)$. Furthermore, if a $\mathsf{SUM} \circ \mathcal{C}$-circuit C always outputs values in the interval $[0,1]$, we say it is a $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$-circuit.

$\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits. Let $\delta \in [0, 0.5)$. A $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit $C\colon \{0,1\}^n \to \{0,1\}$ is defined by a $\mathsf{SUM} \circ \mathcal{C}$-circuit $L\colon \{0,1\}^n \to \mathbb{R}$ satisfying the following promise: for every $x \in \{0,1\}^n$, either $|L(x) - 1| \leq \delta$ or $|L(x)| \leq \delta$. (We stress that this promise is only required over inputs x to the $\mathsf{SUM} \circ \mathcal{C}$-circuit L, and not over all possible input values to the top SUM gate.) We say $C(x) = 1$ if $|L(x) - 1| \leq \delta$ and $C(x) = 0$ otherwise. The sparsity and the complexity of C is defined as the sparsity and the complexity of L, respectively.

For a circuit class $\mathcal{C}$, we use $\mathsf{SUM} \circ \mathcal{C}$, $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$, and $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$ to denote the collection of such circuit families with at most $\mathsf{poly}(n)$ complexity. When $\mathcal{C}$ has a clear notion of complexity, such as circuit size, this also means that the involved $\mathcal{C}$-subcircuits are of polynomial size. In some statements we might refer to classes such as $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}[s]$ to emphasize a specific upper bound s on the complexities of $\mathcal{C}$-subcircuits and of the top gate.

Notation for standard concepts. A $\mathsf{MAJ}\colon \{0,1\}^m \to \{0,1\}$ gate $\mathsf{MAJ}(y_1, \ldots, y_m)$ outputs 1 if and only if $\sum_i y_i \geq m/2$. A $\mathsf{THR}\colon \{0,1\}^m \to \{0,1\}$ gate is described by weights $w_1, \ldots, w_m, \theta \in \mathbb{R}$ and outputs 1 if and only if $\sum_i w_i y_i \geq \theta$.

For a probability distribution $\mathcal{D}$ over $\{0,1\}^n$ and Boolean functions $f, g\colon \{0,1\}^n \to \{0,1\}$, we say that f is γ-approximated by g over $\mathcal{D}$ if $\Pr_{x \sim \mathcal{D}}[f(x) = g(x)] \geq \gamma$. For convenience, circuit lower bounds involving approximations of the form $1/2 + 1/n^{\omega(1)}$ might be informally referred to as *strong average-case lower bounds* or simply *strong correlation bounds*.

Our results refer to non-uniform circuit classes, and we use $\mathcal{C}_1 \circ \mathcal{C}_2$ to refer to circuit families consisting of a top circuit from $\mathcal{C}_1$ composed with bottom circuits from $\mathcal{C}_2$.[1]

We use $\mathcal{U}_n$ to denote the uniform distribution over $\{0,1\}^n$. A distribution $\mathcal{D}$ *ε-fools* a function $f\colon \{0,1\}^n \to \{0,1\}$ if $|\Pr[f(\mathcal{D}) = 1] - \Pr[f(\mathcal{U}_n) = 1]| \leq \varepsilon$. We say that a sequence $G_n\colon \{0,1\}^{\ell(n)} \to \{0,1\}^n$ is an infinitely often PRG against a circuit class $\mathcal{C}$ with error ε (i.o. ε-PRG) and seed length ℓ if G_n is computable in time $2^{O(\ell(n))}$ and for infinitely many values of n, the induced distribution $G_n(\mathcal{U}_{\ell(n)})$ $\varepsilon(n)$-fools each function $f\colon \{0,1\}^n \to \{0,1\}$ in $\mathcal{C}$.

1.2.1 Equivalences for worst-case and strong average-case hardness

Our first contribution is a general result that tightly connects worst-case lower bounds, strong average-case hardness, and pseudorandomness in *restricted computational models*.

[1] As usual, in the case of $\mathcal{C}_2 = \mathsf{ACC}^0$, where $\mathsf{ACC}^0 = \bigcup_{m \in \mathbb{N}} \mathsf{AC}^0[m]$ with m here representing the modulo, we require that each $\mathcal{C}_2$-subcircuit of a circuit D from $\mathcal{C}_1 \circ \mathcal{C}_2$ uses the same fixed m.

Theorem 1 (Non-black-box equivalences for worst-case and strong average-case hardness)**.** *Let $\mathcal{C}$ be a circuit class contained in NC^1 that is closed under negations and under a bottom layer of juntas over $O(1)$ input bits. The following statements are equivalent:*

1. *There is $L \in \mathsf{E}$ such that $L \notin \widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$.*
2. *There is $L \in \mathsf{E}$ and $\delta \geq 1/\mathsf{poly}(n)$ such that $L \notin \widetilde{\mathsf{SUM}}_{\delta} \circ \mathcal{C}$.*
3. *There is $L \in \mathsf{E}$ such that $L \notin \mathsf{MAJ} \circ \mathcal{C}$.*
4. *There is $L \in \mathsf{E}$ such that, for every $k \geq 1$, L cannot be computed by probabilistic $\mathcal{C}$-circuits with error $1/2 - 1/n^k$.*[2]
5. *There is $L \in \mathsf{E}$ and a distribution ensemble $\mathfrak{D}$ such that for every $k \geq 1$, L cannot be $(1/2 + n^{-k})$-approximated by $\mathcal{C}$ under $\mathfrak{D}$.*
6. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, L cannot be $(1/2 + n^{-k})$-approximated by $\mathcal{C}$ under the uniform distribution.*
7. *There is $L \in \mathsf{E}$ that cannot be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$ within ℓ_1 distance $1/3$.*[3]
8. *There is $L \in \mathsf{E}$ and $\delta \geq 1/\mathsf{poly}(n)$ such that L cannot be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$ within ℓ_1 distance δ.*
9. *There is an i.o. ε-PRG G against $\mathcal{C}$ with seed length $n-1$ and error $\varepsilon(n) \leq n^{-\omega(1)}$.*[4]
10. *For each $\gamma > 0$, there is an i.o. ε-PRG against $\mathcal{C}$ with seed length n^γ and $\varepsilon(n) \leq n^{-\omega(1)}$.*

This result can be applied to a variety of natural circuit classes, such as $\mathsf{AC}^0[\oplus]$, ACC^0, and constant-degree polynomial threshold functions (PTFs). We stress that while Theorem 1 requires the circuit class $\mathcal{C}$ to be contained in NC^1, in circuit complexity this is the most interesting case for the result. More precisely, for circuit classes that are above NC^1, it is well known that worst-case hardness for a problem in E can be converted into average-case hardness and PRGs. (Furthermore, NC^1 is closed under a top MAJ or $\widetilde{\mathsf{SUM}}$ gate.) We remark that Theorem 1, with appropriate modifications, can be adapted to other uniform complexity classes, such as $\mathsf{BPE} = \mathsf{BPTIME}[2^{O(n)}]$ and PSPACE. For simplicity, we restrict our discussion to E.

We observe that a connection between worst-case hardness and *weak* average-case hardness for functions in E has been established in [GGH+07], under the assumption that the circuit class $\mathcal{C}$ contains AC^0 and is closed under composition. In contrast to their work, we have a much weaker assumption on $\mathcal{C}$, and our setting of parameters allows us to obtain equivalences to PRGs and to derive consequences that do not follow from their results.

We now highlight three fundamental consequences of Theorem 1. Note that, while our proof employs $\widetilde{\mathsf{SUM}}$ gates in important ways, none of these results refer to such gates.

[2] Following standard terminology, a probabilistic $\mathcal{C}$-circuit F is simply a distribution of $\mathcal{C}$-circuits. We say that F computes a Boolean function g with error ε if for every input x we have $\Pr_F[F(x) \neq g(x)] \leq \varepsilon$.

[3] In other words, there is no family of circuits $F_n \in [0,1]$-$\mathsf{SUM} \circ \mathcal{C}$ such that $\mathsf{E}_{x \sim \{0,1\}^n}[|L(x) - F_n(x)|] \leq 1/3$ for all large n. This notion plays a crucial role in [CLW20] and other related works.

[4] More precisely, for each choice of k, there is an infinite set $S_k \subseteq \mathbb{N}$ such that G fools circuits from $\mathcal{C}[n^k]$ on inputs of length $n \in S_k$ with error $\varepsilon(n) \leq n^{-k}$.

1. Seed reduction for PRGs. Perhaps surprisingly, the equivalence between Items 9 and 10 of Theorem 1 shows the existence of a *generic seed reduction phenomenon* for weak circuit classes. Thus to construct i.o. PRGs of sub-polynomial seed length for a class $\mathcal{C}$ satisfying the conditions of this result it is enough to construct a non-trivial i.o. PRG (i.e. of seed length $\leq n-1$) with small error. In particular, improving the error parameter of the PRG against $\mathsf{AC}^0[\oplus]$ described in [FSUV13] to inverse-super-polynomial would lead to major consequences for $\mathsf{AC}^0[\oplus]$-circuits.

2. Hardness under some distribution implies hardness under the uniform distribution. Theorem 1 also has important implications to our understanding of the average-case hardness of problems in E with respect to weak circuit classes. This is an immediate consequence of Items 5 and 6, which establish the result for strong average-case hardness of the form $1/2 + 1/n^{\omega(1)}$. In Section 3.3, we observe that our techniques can also translate constant-error average-case hardness under an arbitrary distribution to constant-error average-case hardness under the uniform distribution. An interesting application of these results is that the existence of a PRG against $\mathcal{C}$, which was only known to imply hardness under some distribution (see e.g. Section 3 of [Vio09]), also implies hardness with respect to the uniform distribution (which in turn is sufficient to construct PRGs).

3. Equivalence between worst-case and average-case hardness. The well-known Discriminator Lemma from Hajnal et al. [HMP+93] has found numerous applications in circuit complexity lower bounds. It shows that if a Boolean function f cannot be $(1/2 + 1/\mathsf{poly}(n))$-approximated by a class $\mathcal{C}$ then f is not in $\mathsf{MAJ} \circ \mathcal{C}$. In other words, one can lift an average-case lower bound against $\mathcal{C}$ to a worst-case lower bound against the stronger class $\mathsf{MAJ} \circ \mathcal{C}$. Interestingly, the equivalence between Items 3 and 6 in Theorem 1 shows that, for the purpose of proving lower bounds for a problem in E, a worst-case lower bound against $\mathsf{MAJ} \circ \mathcal{C}$ is actually *equivalent* to a strong average-case lower bound against $\mathcal{C}$. To our knowledge, this was previously unknown for weak computational models.[5]

A consequence of Theorem 1 relevant to the study of $\widetilde{\mathsf{SUM}}$ gates is that if $\mathsf{E} \not\subseteq \widetilde{\mathsf{SUM}}_{\delta} \circ \mathcal{C}$ for some $\delta(n) = 1/n^c$ then $\mathsf{E} \not\subseteq \widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$.[6] Another interesting implication is that the average-case lower bounds against $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$ under ℓ_1 distance investigated in [CLW20] are *necessary* and *sufficient* for strong average-case hardness against $\mathcal{C}$.

Next, we discuss some of the techniques behind Theorem 1.

Theorem 1: Techniques. As alluded to above, the proof of Theorem 1 relies on non-black-box hardness amplification techniques explored by Chen and Ren [CR20] and on a careful balance between the *strength* and *weakness* of $\widetilde{\mathsf{SUM}}$ gates. To give some intuition, we discuss the main ingredients behind a more direct proof of the following equivalence, which also explains the assumptions on the circuit class $\mathcal{C}$:

$$\text{Worst-case hardness against } \widetilde{\mathsf{SUM}} \circ \mathcal{C} \iff \text{i.o. PRGs against } \mathcal{C} \text{ with error } \varepsilon = n^{-\omega(1)}.$$

[5] We also remark that it was known [GNW11, Imp95, Kli01] before that for general circuit class $\mathcal{C}$, weak average-case hardness against $\mathsf{MAJ} \circ \mathcal{C}$ implies strong average-case hardness against $\mathcal{C}$.

[6] We note that a simple error amplification technique for $\widetilde{\mathsf{SUM}}$ (Lemma 20) blows up the complexity of the involved $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$-circuits to quasi-polynomial when amplifying from constant-error approximation to inverse polynomial. For this reason, it does not establish this implication.

While it is possible to show that a $\widetilde{\mathsf{SUM}}$ gate can be efficiently simulated by a MAJ gate,[7] the opposite simulation does not hold (e.g. consider approximate degree). In this sense, $\widetilde{\mathsf{SUM}}$ gates are indeed weak. Still, it is possible to show essentially that, for a certain *specific* NC^1-hard problem L contained in P, a $\widetilde{\mathsf{SUM}}$ gate of polynomial complexity can implement a hardness amplification proof: roughly speaking, a weak approximator circuit for L can be transformed into a correct circuit for L by incurring only a top $\widetilde{\mathsf{SUM}}$ gate overhead. This allows us to employ the following win-win analysis. Either the NC^1-hard problem L is $1/2 + n^{-k}$-hard against $\mathcal{C}$ on infinitely many input lengths for every choice of k, in which case an i.o. PRG against $\mathcal{C}$ can be constructed from L using standard techniques under the assumption that $\mathcal{C}$ is closed under bottom layer $O(1)$-juntas, or there is a choice of k such that L can be $1/2+n^{-k}$ approximated by $\mathcal{C}$-circuits on large enough input lengths. The latter implies via the hardness amplification reconstruction routine that $L \in \widetilde{\mathsf{SUM}} \circ \mathcal{C}$, which in turns yields $\mathsf{NC}^1 \subseteq \widetilde{\mathsf{SUM}} \circ \mathcal{C}$ using the NC^1-hardness of L (which in fact admits ultra efficient reductions). Now under our assumption that $\mathcal{C} \subseteq \mathsf{NC}^1$, it is easy to see that $\mathsf{NC}^1 = \widetilde{\mathsf{SUM}} \circ \mathcal{C}$. As a consequence, a worst-case lower bound against $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$ provides a worst-case lower bound against NC^1, and again, PRGs can be constructed from such an assumption via standard methods (since NC^1 admits black-box worst-case to average-case amplification).

For the other direction, we start with an i.o. PRG G against $\mathcal{C}$ that might have a large seed length but guarantees *low error* $\varepsilon(n) = n^{-\omega(1)}$. Here the important insight is that a low error PRG that fools $\mathcal{C}$ also fools linear combinations of functions in $\mathcal{C}$ with bounded coefficients. This implies that G fools $\widetilde{\mathsf{SUM}} \circ C$. Another standard argument shows that from such a PRG one can define a function in E that is worst-case hard against $\widetilde{\mathsf{SUM}} \circ C$.

We stress that two crucial ingredients of our equivalence theorem are the existence of the hard problem L mentioned above and the use of $\widetilde{\mathsf{SUM}}$ gates. The hard language L is actually a pair of problems CMD and DCMD with very useful structural properties (see Section 2.2). They have been explored in a few other works (e.g. [IK02, AIK06, GGH+07, AAW10]), and are tightly connected to *decomposable randomized encodings*, which are well-studied in cryptography (see [App17]). The fruitful interaction between these problems and $\widetilde{\mathsf{SUM}}$ gates was first noticed by [CR20] in the context of the algorithmic method and is a crucial ingredient in their proof that NQP is strongly average-case hard against ACC^0.

While the proof of Theorem 1 avoids the black-box "barrier" and applies to circuit classes that are not assumed to be closed under majority, our techniques come with certain limitations. As a consequence of our indirect analysis via a win-win argument, Theorem 1 does not provide almost-everywhere equivalences for some items and does not scale well to large circuit size bounds above quasi-polynomial. These are important directions for future work.

Applications to ACC^0-circuits lower bound approaches. As a concrete application of Theorem 1 to current frontiers in circuit complexity, we explore its consequences to the average-case complexity of ACC^0. We use our framework to show that existing "combinatorial" approaches to worst-case lower bounds would also provide *strong average-case hardness* against ACC^0. Before

[7] It is possible to approximate all coefficients of the bounded linear sum using sums of powers of 2^i with $i \in \mathbb{Z}$, then multiply the linear sum by an appropriate power of 2 to obtain integer coefficients, and finally simulate the resulting sum by an appropriate THR gate with polynomial weights, which can be translated to a MAJ gate using duplicated input wires and by negating input variables if necessary.

stating this result, we briefly recall some concepts.

Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ be the one-dimensional torus. A *torus polynomial* [BHLR19] (see also [Kri21]) is a real polynomial $p(x_1, \ldots, x_n)$ restricted to the domain $\{0,1\}^n$ and evaluated modulo one.[8] For the purpose of representing the output of a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ as a value in $\mathbb{T}$, we map the output bit $f(x)$ to $f(x)/2$. For $\delta < 1/4$, we say that f is δ-approximated by a degree-d torus polynomial if there is a degree-d real polynomial $p(x_1, \ldots, x_n)$ such that if $f(x) = 1$ then $p(x) - \lfloor p(x) \rfloor \in [1/2 - \delta, 1/2 + \delta]$ and if $f(x) = 0$ then $p(x) - \lfloor p(x) \rfloor \in [0, \delta] \cup [1 - \delta, 1)$. A recent approach proposed by [BHLR19] shows that ACC^0 lower bounds follow from torus polynomial degree lower bounds for approximating a Boolean function.

The number-on-forehead (NOF) multi-party communication model was introduced by [CFL83], and work of [HG91, RW93] show that explicit communication lower bounds in this model (even in the single-round model where all players simultaneously communicate to a referee) imply lower bounds against SYM^+-circuits, which are known to simulate ACC^0 [BT94].

Theorem 2 (Lifting worst-case ACC^0 lower bound approaches to strong correlation bounds)**.** *Consider the following statements:*

1. **Torus Polynomials:** *There is a language $L \in \mathsf{E}$ and a function $\delta(n) \geq 1/\mathsf{poly}(n)$ such that L does not have δ-approximation torus polynomials of degree $\mathsf{polylog}(n)$.*

2. **NOF Protocols:** *There is a language in E that does not admit (single-round) NOF multi-party protocols with $\mathsf{polylog}(n)$ parties of communication cost $\mathsf{polylog}(n)$.*

In each case, if the corresponding statement holds then there is a language in E that cannot be $(1/2 + 1/\mathsf{poly}(n))$-approximated under the uniform distribution by ACC^0.

As a consequence, lower bounds against these models provide i.o. PRGs of sub-polynomial seed length against ACC^0.

Theorem 2: Techniques. It is not hard to adapt classical techniques to show that if a Boolean function can be approximated by torus polynomials of bounded degree, then it can also be computed by NOF protocols of low complexity. For this reason, in order to prove Theorem 2 it is sufficient to obtain average-case hardness against ACC^0 from degree lower bounds for torus polynomials approximating Boolean functions.[9] To achieve this, we refine the argument of [BHLR19] and invoke our framework. In more detail, we show the stronger result that even functions families in $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0$ can be approximated by low-degree torus polynomials. This yields the result using the equivalence between Items 6 and 2 in Theorem 1.

To establish this claim, we make use of low degree "middle-bit polynomials" [GKT92], a sub-class of SYM^+-circuits that is strong enough to simulate ACC^0. By a careful adaptation of the argument of [BHLR19], we are able to show that a linear sum (with bounded coefficients) of middle-bit polynomials with a special structure can be converted into a torus polynomial. The argument is somewhat subtle, and involves the manipulation of universal circuits for depth-d $\mathsf{ACC}^0[s]$ in order to enforce similar parameters for all middle-bit polynomials feeding the top $\widetilde{\mathsf{SUM}}$ gate. The details appear in Section 4.

[8] By a value y (mod 1) we mean its fractional part given by $y - \lfloor y \rfloor$, where the floor function $\lfloor y \rfloor$ denotes the largest integer less than or equal to y. For instance, 1.37 (mod 1) is 0.37 and -2.21 (mod 1) is 0.79.

[9] Alternatively, earlier work on ACC^0 already showed that $\mathsf{MAJ} \circ \mathsf{ACC}^0$-circuits can be simulated by NOF protocols of low communication. Therefore, the NOF protocols part of Theorem 2 follows directly from our Theorem 1.

1.2.2 Lower bounds against circuits with layers of $\widetilde{\mathsf{SUM}}$ and MAJ gates

Observe that Theorem 1 (via Items 1, 2, and 3) establishes the following equivalence: for the purpose of proving circuit lower bounds for a function in E, a top layer MAJ gate is *equivalent* to a top layer $\widetilde{\mathsf{SUM}}$ gate. Given that $\widetilde{\mathsf{SUM}}$ is simpler than MAJ, and lower bounds against $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$ offer a path to correlation bounds and PRGs against $\mathcal{C}$, obtaining a better understanding of $\widetilde{\mathsf{SUM}}$ gates in Boolean circuits might have significant benefits.

In this section, we explore *unconditional* lower bounds against circuits with layers of MAJ and $\widetilde{\mathsf{SUM}}$ gates. Our results are connected to the long-standing problem of showing explicit lower bounds against $\mathsf{THR} \circ \mathsf{THR}$, the class of polynomial-size depth-2 threshold circuits (where size is measured by number of gates). For convenience of the reader, we review below some results related to this frontier.

Threshold circuits. Recall that a threshold gate THR over m input bits is described by weights $w_1, \ldots, w_m, \theta \in \mathbb{R}$. It outputs 1 on an input $y \in \{0,1\}^m$ if and only if $\sum_i w_i y_i \geq \theta$. It is known that every such gate can be implemented with integer weights of magnitude $2^{O(m \log m)}$ (see [Hås94]). In the context of polynomial size circuits, by duplicating input wires a MAJ gate can be equivalently defined as the restriction of a THR gate to polynomially bounded integer weights. It was shown that $\mathsf{MAJ} \circ \mathsf{THR} = \mathsf{MAJ} \circ \mathsf{MAJ}$ and $\mathsf{THR} \circ \mathsf{THR}$ is contained in $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$ [GHR92]. Exponential lower bounds are known against $\mathsf{THR} \circ \mathsf{MAJ}$-circuits [For01], and $\mathsf{THR} \circ \mathsf{MAJ}$ is strictly contained in $\mathsf{THR} \circ \mathsf{THR}$ [CM18]. Recently, [KW16] described a function in P that requires $\mathsf{THR} \circ \mathsf{THR}$-circuits of size (measured by the number of gates) nearly $n^{3/2}$. This is the strongest known lower bound against this class (see their work for extensions to other circuit size measures) for a function in P. It is also known that E^{NP} does not have $n^{2-\varepsilon}$-size $\mathsf{THR} \circ \mathsf{THR}$-circuits for every constant $\varepsilon > 0$ [ACW16, Tam16].

$\mathsf{LTF}^s \circ \mathcal{C}$-circuits: An intermediary class between $\mathsf{MAJ} \circ \mathcal{C}$ and $\mathsf{THR} \circ \mathcal{C}$. In order to make progress toward showing super-polynomial lower bounds against $\mathsf{THR} \circ \mathsf{THR}$-circuits, we study a newly defined gate LTF^s whose power lies between MAJ and THR.[10] Let $\mathsf{SUM}^\infty \circ \mathcal{C}$ be the relaxation of $\mathsf{SUM} \circ \mathcal{C}$ to an *unrestricted* top SUM gate (i.e. the top gate can use arbitrary real coefficients that might not be polynomially bounded). For a given function s and a circuit class $\mathcal{C}$, we say that a function f admits a $\mathsf{LTF}^s \circ \mathcal{C}$-circuit of size S if there is a circuit $D \in \mathsf{SUM}^\infty \circ \mathcal{C}$ such that the following hold: (1) $f(x) = 1$ if and only if $D(x) \geq 0$; (2) $|D(x)| \in (1/s, s)$ for every $x \in \{0,1\}^n$; (3) the total size of the $\mathcal{C}$-subcircuits of D is at most S. Note that unrestricted weights are allowed in the top gate, but we are promised that on each input x the value $D(x)$ is neither too close to 0 nor too large in magnitude.[11]

We are able to extend the algorithmic method [Wil13] to show that #SAT algorithms for a circuit class $\mathcal{C}$ imply worst-case lower bounds against $\mathsf{LTF}^s \circ \mathcal{C}$ and average-case lower bounds against $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$. Let $\mathsf{NQP} = \mathsf{NTIME}[2^{\mathsf{polylog}(n)}]$ be the class of languages computable in non-deterministic quasi-polynomial time. We say that a circuit class $\mathcal{C}$ is *nice* if $\mathcal{C}$ is closed under

[10] LTF denotes linear threshold function, another standard name for THR. We employ both names in this paper to make a clear distinction between the new gates and THR.

[11] Note that we only impose this constraint for each input x of the combined $\mathsf{SUM}^\infty \circ \mathcal{C}$-circuit, and not over all possible input strings for the top gate.

negation, (bottom) projections, and a top AND gate of unbounded fan-in, and in addition $\mathcal{C}$-circuits of size s admit general circuits of depth $O(\log s)$. Examples of nice circuit classes include AC^0, ACC^0, and $\mathsf{AC}^0[\oplus] \circ \mathsf{THR}$.

Theorem 3 (Stronger lower bounds from #SAT algorithms)**.**
Let $\mathcal{C}$ be a nice circuit class. Suppose there is a constant $\varepsilon > 0$ such that, given a $\mathcal{C}$-circuit of size 2^{n^ε} over n input variables, its number of satisfying assignments can be deterministically computed in time 2^{n-n^ε}. Then the following statements hold:

1. *For every constant $k > 0$, NQP does not have $\mathsf{LTF}^{2^{\log^k n}} \circ \mathcal{C}$-circuits of size $2^{\log^k n}$.*
2. *For every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be $(1/2 + 2^{-\log^k n})$-approximated by $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits where both the sparsity of the top SUM-gate and the size of the bottom layer $\mathcal{C}$-circuits are at most $2^{\log^k n}$.*[12]

To our knowledge, these two circuit lower bound consequences are incomparable. By combining Theorem 3 with existing #SAT algorithms for $\mathcal{C} = \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits [Wil18b], we obtain the following unconditional lower bounds.

Corollary 4 (Lower bounds against circuits with $\widetilde{\mathsf{SUM}}$, THR, and MAJ gates)**.**
The following results hold:

1. *For every constant $k > 0$, NQP does not admit $\mathsf{LTF}^{2^{\log^k n}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits of size $2^{\log^k n}$.*
2. *For every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be $(1/2 + 2^{-\log^k n})$-approximated by $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits where the top sum has sparsity $2^{\log^k n}$ and all $\mathsf{ACC}^0 \circ \mathsf{THR}$-subcircuits have size $2^{\log^k n}$.*
3. *For every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be computed by $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits where the top MAJ gate has fan-in $2^{\log^k n}$ and all $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-subcircuits have size and sparsity $2^{\log^k n}$.*

To contrast these results with previous work, we note that [CW19, Theorem 15] gave a *worst-case* lower bound against $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits with any *constant* error δ less than $1/2$. Also, [CR20, Section 5.2] showed a strong average-case lower bound against $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits, where the top sum gate has *zero error* (i.e., $\delta = 0$). Consequently, Corollary 4 Item 2 simultaneously strengthens both results. On the other hand, Corollary 4 Item 3 shows the first lower bound against circuits combining layers of $\widetilde{\mathsf{SUM}}_{1/3}$, MAJ, and THR gates.

Before discussing our techniques in more detail, we mention an open problem and its connection to $\mathsf{THR} \circ \mathsf{THR}$ lower bounds. Recall that this class is contained in $\mathsf{MAJ} \circ \mathsf{MAJ} \circ \mathsf{MAJ}$. In light of the super-polynomial lower bound against $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$ from Corollary 4 Item 3, it would be very interesting to understand the relation between MAJ gates and $\widetilde{\mathsf{SUM}}$ gates appearing in *internal layers* of Boolean circuits. In particular, we note that if MAJ can be simulated by $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathsf{ACC}^0$-circuits of quasi-polynomial size (or THR can be simulated by $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0$-circuits of quasi-polynomial size), then $\mathsf{NQP} \not\subseteq \mathsf{THR} \circ \mathsf{THR}$. On the other hand, if this is not the case, strong average-case lower bounds against ACC^0 follow from Theorem 1.

[12] For the interested reader, we notice that the coefficients of the top $\widetilde{\mathsf{SUM}}$ gate can be unbounded in this lower bound.

Theorem 3 and Corollary 4: Techniques. The proofs of the first two items of Corollary 4 are immediate from the corresponding items of Theorem 3 via the #SAT algorithm for $\mathcal{C} = \mathsf{ACC}^0 \circ \mathsf{THR}$ given by [Wil18b]. On the other hand, Item 3 of Corollary 4 can be established in different ways. The first proof is just a standard application of the Discriminator Lemma [HMP+93] together with the lower bound from Item 2. A second proof follows from Item 1, via a simulation of a $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit of quasi-polynomial complexity by a $\mathsf{LTF}^{2^{\log^k n}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuit of size size $2^{\log^k n}$, for some constant k. This can be done by first reducing the error δ of each $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-subcircuit (see Lemma 20), then rewriting the corresponding $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\varepsilon$ top layers as an LTF^s gate via an appropriate collapse. We omit the details.

The proofs of Items 1 and 2 of Theorem 3 are essentially independent. We discuss each of them next, starting with Item 1.

An extension of the algorithmic method [Wil13] obtained by [MW20] shows that SAT algorithms for a circuit class $\mathcal{C}$ of sub-exponential size circuits (satisfying minor closure conditions) that run in time 2^{n-n^ε} imply that $\mathsf{NQP} \not\subseteq \mathcal{C}$. In a more recent work that builds on [Wil18a], [CW19] established (in particular) that #SAT algorithms of similar running time provide the stronger lower bound $\mathsf{NQP} \not\subseteq \widetilde{\mathsf{SUM}} \circ \mathcal{C}$. Our proof of Item 1 of Theorem 3 relies on the latter result and on a win-win argument inspired by [CR20]. In more detail, and oversimplifying a bit, we argue that if a special NC^1-hard problem L (contained in NQP) is not in $\mathsf{LTF}^{2^{\log^k n}} \circ \mathcal{C}$, then we are done. Otherwise, we explore LTF^s gates and the special form of the NC^1-hardness of L to show that NC^1 can be simulated by $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$-circuits of quasi-polynomial complexity. Given this lemma and the corresponding simulation, we can reduce the derivation of the desired lower bound to previous work, i.e., we invoke the aforementioned connection between #SAT algorithms and lower bounds against $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$. This provides a language in NQP that is not in $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$ of complexity $2^{\log^\ell n}$, where $\ell = \ell(k)$ is large enough. Now by simulating $\mathsf{LTF}^{2^{\log^k n}} \circ \mathcal{C}$-circuits using quasi-polynomial size Boolean formulas, and using the collapse of NC^1 to quasi-polynomial size $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$, it is possible to argue that L is also hard against $\mathsf{LTF}^{2^{\log^k n}} \circ \mathcal{C}$.

The proof of Item 2 of Theorem 3 shares some similarities with the argument above, but the technical details are different. From a high-level perspective, we also employ a win-win argument, though this time it is based on the *average-case* complexity of the language L mentioned above. Moreover, we cannot rely on previous connections between #SAT algorithms and lower bounds in a *black-box* way. Given that explaining the relevant details would be fairly technical, we refer the interested reader to Section 5.3. We mention that a conceptual contribution is that while our proof of Theorem 3 Part 2 follows the strategy of previous works, such as [Che19, CW19, CR20], on obtaining lower bounds from meta-algorithms, it does not use PCPs of proximity (PCPP), which was a key ingredient in the proofs of those works. For this, we rely in part on a PCP stated in [Vio20], combined with other ideas.

Organization of this paper

In Section 2 we introduce the necessary technical preliminaries for proving our results. In Section 3 we prove our main equivalence result (Theorem 1). In Section 4 we present consequences of our equivalence theorem to previous approaches to ACC^0 lower bounds (Theorem 2). Finally, in Section 5 we prove our worst-case lower bound against $\mathsf{LTF}^{\mathsf{quasi\text{-}poly}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits and strong average-case lower bound against $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits (Theorem 3 and Corollary 4).

2 Preliminaries

2.1 Notation

We use $\mathbb{N}$ to denote the set of all non-negative integers and $\mathbb{N}_{\geq 1}$ to denote $\mathbb{N} \setminus \{0\}$. For every $n \in \mathbb{N}_{\geq 1}$, we let $\mathcal{U}_n$ denote the uniform distribution over $\{0,1\}^n$. For convenience, in some settings a Boolean function $f\colon \{0,1\}^n \to \{0,1\}$ will be viewed as a function with output in $\{-1,1\}$, where -1 and 1 are interpreted as True and False, respectively.

For a predicate $P(x)$, we use $\mathbb{1}_{P(x)}$ to denote its corresponding Boolean value on x. That is, $\mathbb{1}_{P(x)} = 1$ if $P(x)$ is true, and 0 otherwise. For a real v, we define $\mathrm{sign}(v) := (-1) \cdot \mathbb{1}_{v<0} + 1 \cdot \mathbb{1}_{v \geq 0}$.

For two strings $\alpha, \beta \in \{0,1\}^*$, we write $\alpha \circ \beta$ to denote the concatenation of α and β.

A projection of a function $f(x_1, \ldots, x_n)$ is a function $g(y_1, \ldots, y_m)$ with a projection mapping $P\colon \{0,1\}^m \to \{0,1\}^n$ such that $g(y_1, \ldots, y_m) = f(P(y_1, \ldots, y_m))$. By "projection" we mean that each output bit of $P(y_1, \ldots, y_m)$ is either an input bit y_i, its negation, or a constant.

Let a be a positive integer. For an arbitrary $\ell \geq 1$ and a function $h\colon \{0,1\}^\ell \to \{0,1\}$, we say that $h \in \mathsf{JUNTA}_a$ if the output of h depends on at most a input coordinates.

2.2 A $\oplus$L-complete problem with good properties

The existence of $\oplus$L-complete problems with good reducibility properties will be important for us. (Recall that $\oplus$L is the class of problems solvable by a nondeterministic logspace Turing machine that accepts the input if the number of accepting paths is odd.) We define the following two problems, called Connected Matrix Determinant (CMD) and Decomposed Connected Matrix Determinant (DCMD):

Definition 5. *An instance of* CMD *is an $n \times n$ matrix over $\mathbb{F}_2$ where the main diagonal and above may contain either 0 or 1, the second diagonal (i.e. the one below the main diagonal) contains 1, and other entries are 0. In other words, the matrix is of the following form (where $*$ represents any element in $\mathbb{F}_2$):*

$$\begin{pmatrix} * & * & * & \cdots & * & * \\ 1 & * & * & \cdots & * & * \\ 0 & 1 & * & \cdots & * & * \\ 0 & 0 & 1 & \cdots & * & * \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & * \end{pmatrix}.$$

The instance is an $(n(n+1)/2)$-bit string specifying elements on and above the main diagonal. We define $x \in$ CMD if and only if the determinant (over $\mathbb{F}_2$) of the matrix corresponding to x is 1.

An instance of DCMD *is a string of length $n^3(n+1)/2$. For an input x,* DCMD(x) *is computed as follows: we partition x into blocks of length n^2, let y_i $(1 \leq i \leq n(n+1)/2)$ be the parity of the i-th block, and define* DCMD$(x) :=$ CMD$(y_1 \circ y_2 \circ \cdots \circ y_{n(n+1)/2})$.

The precise definitions of CMD and DCMD are not important here, as we only need the following two important results about them.

Theorem 6 ([AIK06, GGH+07])**.** *There is a function $P : \{0,1\}^{n(n+1)/2} \times \{0,1\}^{O(n^4)} \to \{0,1\}^{n^3(n+1)/2}$ such that the following hold.*

- *For any input* $x \in \{0,1\}^{n(n+1)/2}$*, the random variable* $P(x, \mathcal{U}_{O(n^4)})$ *is uniformly distributed in* $\{0,1\}^{n^3(n+1)/2}$*.*
- *For any* $x \in \{0,1\}^{n(n+1)/2}$ *and* $r \in \{0,1\}^{O(n^4)}$*, let* $P(x,r) = y$*, then* $\mathsf{CMD}(x) = \mathsf{DCMD}(y) \oplus r_0$*, where* r_0 *is the first bit of* r*.*
- *For each fixed randomness* r*,* $P(x,r)$ *is a projection over* x*, computable in polynomial time given* r*.*

Theorem 7 ([IK02])**.** CMD *is* $\oplus\mathsf{L}$*-complete under projections.*

Observe that if CMD is in a circuit class $\mathcal{C}$ closed under projections then all problems in (non-uniform) NC^1 are also in $\mathcal{C}$, given that the problem of evaluating an input Boolean formula is solvable with logarithmic space.

We refer the reader to the full version of [CR20] for a self-contained exposition of these problems and their relevant properties, including pointers to related work.

3 Equivalences for worst-case and strong average-case lower bounds

In this section, we prove our equivalence results for worst-case hardness, strong average-case hardness and pseudorandomness. We start with some useful facts in Section 3.1, then prove Theorem 1 in Section 3.2. In Section 3.3, we further investigate equivalences in the constant-error regime and their consequences.

3.1 Preliminaries

Notation. For a circuit class $\mathcal{C}$ and $s \geq 1$, we use $\widetilde{\mathsf{SUM}} \circ \mathcal{C}[s]$ to denote the class of $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$-circuits where the top SUM gate has complexity at most s and the bottom layer $\mathcal{C}$-circuits have size at most s.

For a function $f\colon \{0,1\}^n \to \{0,1\}$, we let $f_{\pm}\colon \{0,1\}^n \to \{-1,1\}$ be the $\{-1,1\}$-version of f where we map the output of f from 0 to 1 and 1 to -1. Also, for a circuit class $\mathcal{C}$ where the circuits in $\mathcal{C}$ output values in $\{0,1\}$, we denote by $\mathcal{C}_{\pm}$ the $\{-1,1\}$-version of $\mathcal{C}$ where the circuits in $\mathcal{C}_{\pm}$ output values in $\{-1,1\}$.

Pseudorandomness. We need the following Hardness vs. Randomness framework for constructing PRGs.

Lemma 8 (Hardness vs. Randomness [NW94], see also [CR20, Appendix E.3] for the proof)**.** *There is a function* $G\colon \{0,1\}^* \times \{0,1\}^* \to \{0,1\}^*$ *such that the following holds. Let* n, ℓ, a *be integers such that* $a \leq \ell$*, and* $t = O\big(\ell^2 \cdot n^{1/a}/a\big)$*. Let* $\mathcal{C}$ *be a circuit class closed under negation. For any function* $Y\colon \{0,1\}^\ell \to \{0,1\}$ *represented as a length-*2^ℓ *truth table, if* Y *cannot be* $(1/2 + \varepsilon/n)$*-approximated by* $\mathcal{C} \circ \mathsf{JUNTA}_a$*-circuits where the top circuit has size* S*, then for every circuit* $C \in \mathcal{C}$ *of size* S*,*

$$\left| \mathop{\mathbf{Pr}}_{z \sim \{0,1\}^t}[C(G(Y,z)) = 1] - \mathop{\mathbf{Pr}}_{x \sim \{0,1\}^n}[C(x) = 1] \right| \leq \varepsilon.$$

Moreover, the function G *is computable in* $\mathsf{poly}\big(n, 2^t\big)$ *time.*

The following simple fact says PRGs imply worst-case hardness.

Proposition 9 (Worst-case hardness from PRGs)**.** *Let $\mathcal{F}$ be a class of functions. If there is an i.o. ε-PRG $G\colon \{0,1\}^r \to \{0,1\}^n$ with seed length $r(n)$ against $\mathcal{F}_n$, where $\varepsilon < 1 - 2^{r(n)-n}$, then there is a language $L \in E$ such that L cannot be computed by $\mathcal{F}$.*

Proof. Let $G\colon \{0,1\}^r \to \{0,1\}^n$ be an i.o. ε-PRG against $\mathcal{F}$, where $\varepsilon < 1 - 2^{r-n}$. We define $L_n := \{x : \exists y \in \{0,1\}^r \text{ s.t. } G(y) = x\}$. For the sake of contradiction, suppose for every n there is some function $f \in \mathcal{F}_n$ that computes L_n. On the one hand, we have

$$\mathop{\mathbf{E}}_{x\in\{0,1\}^n}[f(x)] = \mathop{\mathbf{E}}_{x\in\{0,1\}^n}[L_n(x)] \leq 2^{r-n}.$$

On the other hand, we have

$$\mathop{\mathbf{E}}_{y\in\{0,1\}^r}[f(G(y))] = 1.$$

Therefore, we have

$$\left|\mathop{\mathbf{E}}_{y\in\{0,1\}^r}[f(G(y))] - \mathop{\mathbf{E}}_{x\in\{0,1\}^n}[f(x)]\right| \geq 1 - 2^{r-n} > \varepsilon,$$

which contradicts the security of G. □

Hardness amplification. The following result allows us to amplify hardness against NC^1.

Lemma 10 (Hardness amplification against NC^1, see e.g. [STV01, GGH+07])**.** *Suppose there is a language $L \in \mathsf{E}$ such that $L \notin \mathsf{NC}^1$. Then there is a language $L' \in \mathsf{E}$ such that for every constant $k \geq 1$, L' cannot be $(1/2 + 1/n^k)$-approximated by formulas of size n^k.*

The following notion of ℓ_1-approximation by SUM-circuits plays a crucial role in some recent results on average-case lower bounds via the algorithmic method (e.g. [CLW20, CL21, HV21]).

Definition 11 (ℓ_1-approximation by SUM-circuits)**.** *Let $\delta \in (0,1)$ and let $\mathcal{C}$ be a circuit class. We say that a function $f\colon \{0,1\}^n \to \{0,1\}$ is approximated by a $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$-circuit C within ℓ_1 distance δ if*

$$\mathop{\mathbf{E}}_{x\sim\mathcal{U}_n}[\,|f(x) - C(x)|\,] \leq \delta.$$

For functions $f, g\colon \{0,1\}^n \to \mathbb{R}$, we let $\langle f, g\rangle := \mathbf{E}_{x\in\{0,1\}^n}[f(x)\cdot g(x)]$.

Proposition 12. *Let $\delta \in (0,1)$, $f\colon \{0,1\}^n \to \{0,1\}$, and $\mathcal{C}$ be a circuit class.*

1. *If f can be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$-circuits of complexity s within ℓ_1 distance δ, then there is a $\mathsf{SUM} \circ \mathcal{C}_{\pm}$-circuit C of complexity $O(s)$ such that $\|C\|_\infty \leq 1$ and $\langle f_{\pm}, C\rangle \geq 1 - 2\delta$.*

2. *If there is a $\mathsf{SUM} \circ \mathcal{C}_{\pm}$-circuit C of complexity s such that $\|C\|_\infty \leq 1$ and $\langle f_{\pm}, C\rangle \geq 1 - 2\delta$, then f can be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$-circuits of complexity $O(s)$ within ℓ_1 distance δ.*

Proof. We first show Item 1. Let C_0 be a $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}$-circuit of complexity s that approximates f within ℓ_1 distance δ. Consider the following

$$C(x) := 1 - 2C_0(x).$$

Note that C can be written as a $\mathsf{SUM} \circ \mathcal{C}_{\pm}$-circuit of complexity $O(s)$ and that $\|C\|_\infty \leq 1$. Also, since $f_{\pm}(x) \in \{-1, 1\}$, for every x we have

$$|f_{\pm}(x) - C(x)| + f_{\pm}(x) \cdot C(x) = 1,$$

which gives

$$\begin{aligned}\langle f_{\pm}, C\rangle &= 1 - \|f_{\pm} - C\|_1 \\ &= 1 - \|(1-2f) - (1-2C_0)\|_1 \\ &= 1 - 2\|f - C_0\|_1 \\ &\geq 1 - 2\delta.\end{aligned}$$

The proof above can be easily adapted to show Item 2 and we omit the details here. □

Given a set X and a Boolean function $f\colon X \to \{-1,1\}$, for and integer $t \geq 1$ and $X^t = X \times \ldots \times X$ (t times) we let $f^{\oplus t}\colon X^t \to \{-1,1\}$ be the Boolean function defined as $f^{\oplus t}(x_1, \ldots, x_t) \coloneqq \prod_{i\in[t]} f(x_i)$. We will need the following XOR lemma from [CLW20].

Theorem 13 ([Lev87] and [CLW20, Lemma 3.8], see also [CL21, Lemma 1.7])**.** *Let $\mathcal{F}$ be a class of Boolean functions that is closed under negation and restriction. For every $\delta, \varepsilon \in (0,1)$ and every function $f\colon \{0,1\}^n \to \{-1,1\}$, if*

$$\langle f, C\rangle \leq 1 - \delta$$

for every $\mathsf{SUM} \circ \mathcal{F}$-circuit C where the top SUM has complexity $10 \cdot n/\varepsilon^2$ and $\|C\|_\infty \leq 1$, then

$$\langle f^{\oplus t}, D\rangle \leq (1-\delta)^t + \varepsilon/\delta$$

for any Boolean function $D\colon \{0,1\}^{tn} \to \{-1,1\}$ in $\mathcal{F}$.

3.2 Proof of Theorem 1

In this subsection, we prove Theorem 1 (restated below).

Reminder of Theorem 1. *Let $\mathcal{C}$ be a circuit class that satisfies the following:*

- *$\mathcal{C}$ is closed under negation and projection.*
- *$\mathcal{C}$ is closed under a bottom layer of juntas over $O(1)$ input bits. That is*

$$\bigcup_{k\geq 1} \mathcal{C}[n^k] \circ \mathsf{JUNTA}_k \subseteq \bigcup_{k\geq 1} \mathcal{C}[n^k].$$

- $\bigcup_{k\geq 1} \mathcal{C}[n^k] \subseteq \mathsf{NC}^1$.

Then the following statements are equivalent:

1. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, $L \notin \widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}[n^k]$.*
2. *There is $L \in \mathsf{E}$ and $\delta \geq 1/\mathsf{poly}(n)$ such that for every $k \geq 1$, $L \notin \widetilde{\mathsf{SUM}}_{\delta} \circ \mathcal{C}[n^k]$.*

3. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, $L \notin \mathsf{MAJ} \circ \mathcal{C}[n^k]$.*

4. *There is $L \in \mathsf{E}$ such that, for every $k \geq 1$, L cannot be computed by a probabilistic $\mathcal{C}[n^k]$-circuit with error $1/2 - 1/n^k$.*

5. *There is $L \in \mathsf{E}$ and a distribution $\mathcal{D}$ such that for every $k \geq 1$, L cannot be $(1/2 + n^{-k})$-approximated by $\mathcal{C}[n^k]$ under $\mathcal{D}$.*

6. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, L cannot be $(1/2 + n^{-k})$-approximated by $\mathcal{C}[n^k]$ under the uniform distribution.*

7. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, L cannot be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}[n^k]$ within ℓ_1 distance $1/3$.*

8. *There is $L \in \mathsf{E}$ and $\delta \geq 1/\mathsf{poly}(n)$ such that for every $k \geq 1$, L cannot be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}[n^k]$ within ℓ_1 distance δ.*

9. *There is an i.o. ε-PRG G against $\mathcal{C}$ with seed length $n - 1$ and error $\varepsilon(n) \leq n^{-\omega(1)}$.*
   In other words, for each choice of k, there is an infinite set $S_k \subseteq \mathbb{N}$ such that G fools circuits from $\mathcal{C}[n^k]$ on inputs of length $n \in S_k$ with error $\varepsilon(n) \leq n^{-k}$.

10. *For every $\gamma > 0$, there is an i.o. ε-PRG against $\mathcal{C}$ with seed length n^γ and $\varepsilon(n) \leq n^{-\omega(1)}$.*

Proof.

Proof outline. We will first show Item 2 ⇒ Item 6 ⇒ Item 10 ⇒ Item 1 ⇒ Item 2, establishing the equivalence of Items 1, 2, 6 and 10. We then show Item 6 ⇒ Item 5 ⇒ Item 4 ⇒ Item 1, which adds Items 4 and 5 to the list of equivalent items. Next, we show Item 6 ⇒ Item 3 ⇒ Item 4, which adds Item 3, and Item 10 ⇒ Item 9 ⇒ Item 2, which adds Item 9. Finally, we show Item 6 ⇒ Item 7 ⇒ Item 8 ⇒ Item 6, adding Items 7 and 8 to the list and completing the proof. See also Figure 1.

Item 2 ⇒ Item 6. We consider two cases. If DCMD cannot be $\left(1/2 + 1/n^k\right)$-approximated by $\mathcal{C}[n^k]$ for every $k \geq 1$ under the uniform distribution, then we are done.

Now consider the case that there is some $k \geq 1$ such that DCMD can be $\left(1/2 + 1/n^k\right)$-approximated by $\mathcal{C}[n^k]$. By the random self-reducibility of $\mathsf{DCMD}/\mathsf{CMD}$ (see Theorem 6 and also [CR20, Section 3]), for any $\delta = 1/\mathsf{poly}(n)$, CMD can be computed by a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit where the top SUM-gate has polynomial complexity and the bottom-layer $\mathcal{C}$-circuits have polynomial size. By Theorem 7, for every polynomial-size parity branching program B, there is a projection $p\colon \{0,1\}^n \to \{0,1\}^{n^{O(1)}}$ such that for every $x \in \{0,1\}^n$, $B(x) = \mathsf{CMD}(p(x))$. Since $\mathcal{C}$ is closed under projection, this means that every polynomial-size parity branching program has a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit of polynomial complexity and size, which then implies that every function in NC^1 also has such a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit. On the other hand, by Item 2, there is a function $L \in \mathsf{E}$ that has no $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit of polynomial complexity and size, so L is not in NC^1. Using hardness amplification against NC^1 (Lemma 10), it follows that there is a function in E that is strongly average-case hard against NC^1, which by assumption contains polynomial-size $\mathcal{C}$-circuits.

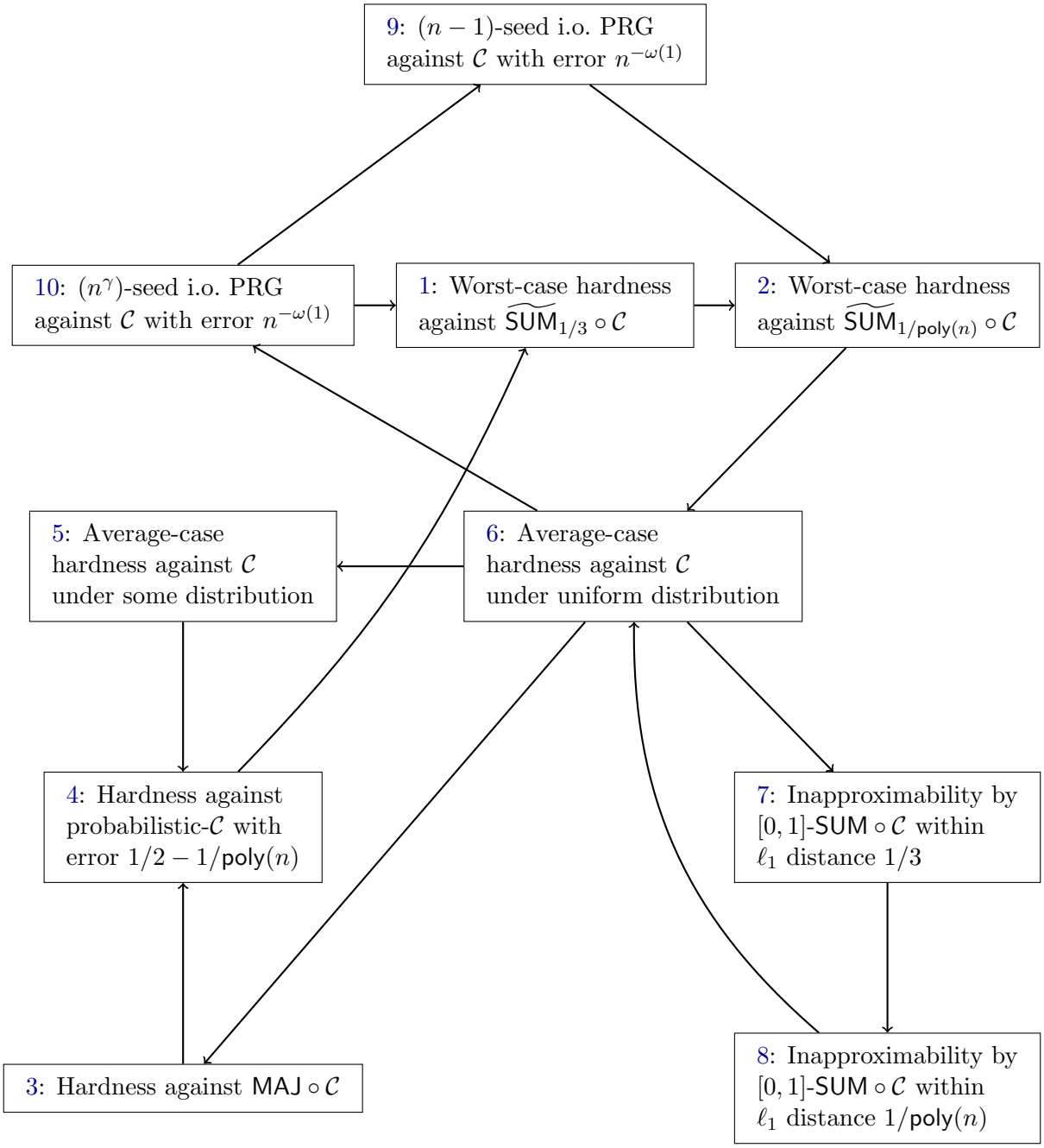

Figure 1: Equivalences in Theorem 1, with arrows indicating direct implications that we prove in order to establish them.

Item 6 ⇒ Item 10. We construct the PRG using the hardness vs. randomness framework. Consider Lemma 8 with the following setting of parameters: $a := 2/\gamma$ and $\ell := n^{\gamma/4}$. Let $G_{L_\ell}\colon \{0,1\}^t \to \{0,1\}^n$ be the PRG defined as $G_{L_\ell}(z) := G(L_\ell, z)$, where $L \in \mathsf{E}$ is the language from Item 6. Note that the seed length t is at most $O\left(\ell^2 \cdot n^{1/a}/a\right) \leq n^\gamma$ and G_{L_ℓ} can be computed in time $\mathsf{poly}(n, 2^t) = 2^{O(n^\gamma)}$. Let $k \geq 1$ be any constant and consider any ℓ-variate $\mathcal{C} \circ \mathsf{JUNTA}_a$-circuit C where the top circuit has size $n^k = \ell^{4k/\gamma}$. Since $\mathcal{C}$ is closed under a bottom layer of juntas, we have that $C \in \mathcal{C}\left[\ell^{k'}\right]$ for some large enough $k' > 4k/\gamma$. Also, let $\varepsilon = 1/n^k$, which implies $\varepsilon/n = 1/n^{k+1} = 1/\ell^{4(k+1)/\gamma} \geq 1/\ell^{k'}$. From Item 6, we have that L_ℓ cannot be $\left(1/2 + 1/\ell^{k'}\right)$-approximated by any circuit from $\mathcal{C}[\ell^{k'}]$, for infinitely many values of ℓ. Then by Lemma 8, we conclude that G_{L_ℓ} $\left(1/n^k\right)$-fools any circuit from $\mathcal{C}[n^k]$, for infinitely many values of n.

Item 10 ⇒ Item 1. Let $G\colon \{0,1\}^r \to \{0,1\}^n$ be an i.o. PRG as in Item 10, where $r \leq n-2$. That is, for each choice of k', G fools circuits from $\mathcal{C}[n^{k'}]$ on input length n with error $\varepsilon(n) \leq n^{-k'}$, for infinitely many values of n.

Let $k \geq 1$ and let $C \in \widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}[n^k]$. By Proposition 9, it suffices to show that G is an i.o. $\left(< \frac{3}{4}\right)$-PRG against C. Let $\tilde{C}$ be the corresponding linear sum for C. That is,

$$\tilde{C}(x) := \sum_i \alpha_i \cdot C_i(x),$$

where $C_i \in \mathcal{C}[n^k] \subseteq \mathcal{C}[n^{k':=k+1}]$ and $\sum_i |\alpha_i| \leq n^k$. Since $\tilde{C}$ $(1/3)$-approximates C in a pointwise manner, we have

$$\left|\mathbf{E}[C(\mathcal{U})] - \mathbf{E}[\tilde{C}(\mathcal{U})]\right| \leq 1/3 \text{ and } \left|\mathbf{E}[C(G)] - \mathbf{E}[\tilde{C}(G)]\right| \leq 1/3.$$

Therefore, if we can show that

$$\left|\mathbf{E}[\tilde{C}(\mathcal{U})] - \mathbf{E}[\tilde{C}(G)]\right| \leq \delta,$$

for some $\delta < 1/12$ (infinitely often), then G δ'-fools C (infinitely often), where $\delta' = 2/3 + \delta < 3/4$. We have

$$\begin{aligned}
\left|\mathbf{E}[\tilde{C}(\mathcal{U})] - \mathbf{E}[\tilde{C}(G)]\right| &= \left|\mathbf{E}\left[\sum_i \alpha_i \cdot C_i(\mathcal{U})\right] - \mathbf{E}\left[\sum_i \alpha_i \cdot C_i(G)\right]\right| \\
&= \left|\sum_i \alpha_i \cdot \mathbf{E}[C_i(\mathcal{U})] - \mathbf{E}[C_i(G)]\right| \\
&\leq \max_i |\mathbf{E}[C_i(\mathcal{U})] - \mathbf{E}[C_i(G)]| \cdot \sum_i |\alpha_i| \\
&\leq n^{-k'} \cdot n^k \leq 1/n,
\end{aligned}$$

as desired.

Item 1 ⇒ Item 2. This implication is straightforward.

Item 6 ⇒ Item 5 ⇒ Item 4. The first implication is obvious. The contrapositive of the second implication follows from an averaging argument.

Item 4 ⇒ Item 1. It suffices to show that for every $k \geq 1$, every function in $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}[n^k]$ has a probabilistic $\mathcal{C}[n^k]$-circuit with error $1/2 - 1/n^{O(k)}$.

For the simplicity of presentation, we will consider Boolean functions that take inputs from $\{0,1\}^n$ and output values in $\{-1,1\}$. Let $f_\pm \colon \{0,1\}^n \to \{-1,1\} \in \widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}_\pm[n^k]$. Then there is a linear sum of $\mathcal{C}_\pm[n^k]$-circuits

$$f_1(x) := \sum_i \alpha_i \cdot \left(\frac{1 - C_i(x)}{2}\right),$$

where $C_i \colon \{0,1\}^n \to \{-1,1\} \in \mathcal{C}_\pm[n^k]$ and $\sum_i |\alpha_i| \leq n^k$, such that

- if $f_\pm(x) = 1$, then $f_1(x) \leq 1/3$, and
- if $f_\pm(x) = -1$, then $f_1(x) \geq 2/3$.

Next, let

$$f_2(x) := 1/2 - f_1(x).$$

It is easy to see that

- if $f_\pm(x) = 1$, then $f_2(x) \geq 1/6$, and
- if $f_\pm(x) = -1$, then $f_2(x) \leq -1/6$.

Now note that since $\mathcal{C}_\pm$ is closed under negation, f_2 can be written as

$$f_2(x) := \sum_j \beta_j \cdot D_j(x),$$

where for each j, $D_j \colon \{0,1\}^n \to \{-1,1\} \in \mathcal{C}_\pm[n^k]$, $\beta_j \geq 0$, and $T := \sum_j \beta_j \leq n^{O(k)}$. Finally, let

$$f_3(x) := \frac{f_2(x)}{T}.$$

Let $\mathcal{D}$ be the probabilistic $\mathcal{C}_\pm[n^k]$-circuit where D_j is sampled with probability β_j/T. Then for every x we have $\mathbf{E}_{\mathcal{D}}[\mathcal{D}(x)] = f_3(x)$. Moreover, if $f_\pm(x) = 1$, then

$$\begin{aligned}
\frac{1}{6T} &\leq \mathop{\mathbf{E}}_{\mathcal{D}}[\mathcal{D}(x)] \\
&= \mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = 1] - \mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = -1] \\
&= \mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = 1] - (1 - \mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = 1]) \\
&= 2 \cdot \mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = 1] - 1,
\end{aligned}$$

which implies

$$\mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = 1] \geq \frac{1}{2} + \frac{1}{12T}.$$

Similarly, we can show that if $f_\pm(x) = -1$, then

$$\mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = -1] \geq \frac{1}{2} + \frac{1}{12T}.$$

Therefore, $\mathcal{D}$ is a probabilistic $\mathcal{C}_\pm[n^k]$-circuit for $f_\pm$ with error $1/2 - 1/n^{O(k)}$.

Item 6 ⇒ Item 3. This follows from the standard Discriminator Lemma [HMP+93].

Item 3 ⇒ Item 4. We will show that for every $k \geq 1$, every function that has a probabilistic $\mathcal{C}[n^k]$-circuit with error $1/2 - 1/n^k$ is contained in $\mathsf{MAJ}_{n^{O(k)}} \circ \mathcal{C}[n^{O(k)}]$.

Let $f\colon \{0,1\}^n \to \{0,1\}$ and $\mathcal{D}$ be the probabilistic $\mathcal{C}[n^k]$-circuit for f with error $1/2 - 1/n^k$. That is, for every x,

$$\mathop{\mathbf{Pr}}_{\mathcal{D}}[\mathcal{D}(x) = f(x)] \geq 1/2 + 1/n^k.$$

By the Chernoff bound, if we sample $t := O(n^{2k} \cdot n)$ circuits $C_1, \ldots, C_t$ from $\mathcal{D}$, then

$$\mathop{\mathbf{Pr}}_{C_1,\ldots,C_t \sim \mathcal{D}} \left[\mathop{\mathbf{Pr}}_{i \in [t]} [C_i(x) = f(x)] \geq 1/2 + 1/(2n^k) \right] \geq 1 - 2^{-2n}.$$

By a union bound over $x \in \{0,1\}^n$, there exist t circuits $C_1, \ldots, C_t$ such that for every x,

$$\mathop{\mathbf{Pr}}_{i \in [t]} [C_i(x) = f(x)] \geq 1/2 + 1/(2n^k).$$

Therefore, by taking the majority of these t circuits, we obtain a $\mathsf{MAJ}_{n^{O(k)}} \circ \mathcal{C}[n^{O(k)}]$-circuit that computes f.

Item 10 ⇒ Item 9 ⇒ Item 2. This first implication is obvious. The proof of the second implication is essentially the same as that of "Item 10 ⇒ Item 1". From Item 9, we get an i.o. PRG with seed length $n-1$ that $\left(< \frac{1}{2}\right)$-fools $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits for some $\delta = 1/\mathsf{poly}(n)$, which by Proposition 9 implies Item 2. We omit the details here.

Item 6 ⇒ Item 7. Let $L\colon \{0,1\}^* \to \{0,1\}$ be the language from Item 6. For the sake of contradiction, suppose there is a $k \geq 1$ such that L can be approximated by $[0,1]$-$\mathsf{SUM} \circ \mathcal{C}[n^k]$-circuits within ℓ_1 distance $1/3$. Then by Item 1 of Proposition 12, we have that for every n, there is a $\mathsf{SUM} \circ \mathcal{C}_{\pm}[O\left(n^k\right)]$-circuit C such that $\|C\|_\infty \leq 1$ and

$$\langle (L_{\pm})_n, C \rangle \geq 1/3.$$

Suppose

$$C(x) := \sum_i |\alpha_i| \cdot C_i(x),$$

where $C_i \in \mathcal{C}_{\pm}[O\left(n^k\right)]$ and $\sum_i |\alpha_i| \leq O\left(n^k\right)$. Then

$$\begin{aligned}
1/3 &\leq \left\langle (L_{\pm})_n, \sum_i \alpha_i \cdot C_i \right\rangle \\
&= \sum_i \alpha_i \cdot \langle (L_{\pm})_n, C_i \rangle \\
&\leq \sum_i |\alpha_i| \cdot \langle (L_{\pm})_n, C_i \rangle \\
&\leq \max_i \langle (L_{\pm})_n, C_i \rangle \cdot \sum_i |\alpha_i| \\
&\leq \max_i \langle (L_{\pm})_n, C_i \rangle \cdot O\left(n^k\right),
\end{aligned}$$

which implies that there exists some i such that

$$\langle (L_{\pm})_n, C_i \rangle \geq \frac{1}{O(n^k)}.$$

This contradicts Item 6.

Item 7 $\Rightarrow$ Item 8. This implication is obvious.

Item 8 $\Rightarrow$ Item 6. By Item 2 of Proposition 12, we have that Item 8 implies that there is a language $L\colon \{0,1\}^* \to \{-1,1\}$ in E and $\delta = 1/\ell^b$, where $b \geq 1$ is a constant, such that for every $k' \geq 1$, on infinitely many input lengths there is no $\mathsf{SUM} \circ \mathcal{C}_{\pm}[\ell^{k'}]$-circuit C with $\|C\|_{\infty} \leq 1$ such that

$$\langle L_\ell, C \rangle \leq 1 - 2\delta. \tag{1}$$

Now consider the following language $L'\colon \{0,1\}^* \to \{-1,1\}$: on input x of length n, let ℓ be the largest integer such that $\ell \cdot \ell^b \log^2(\ell) \leq n$ and view the input as $x = (x_1, \ldots, x_t, y)$, where $t := \ell^b \log^2(\ell)$ and $x_i \in \{0,1\}^\ell$ for $i \in [t]$. Then let

$$L'(x) := \prod_{i \in [t]} L(x_i).$$

Note that for large enough n we have

$$n < 2\ell \cdot t < \ell^{b+2}.$$

We claim that L' is strongly average-case hard against $\mathcal{C}_{\pm}$-circuits. For the sake of contradiction, suppose there is $k \geq 1$ and an n-variate circuit $C' \in \mathcal{C}_{\pm}[n^k]$ such that, for all large enough n,

$$\langle L'_n, C' \rangle > \frac{1}{n^k}.$$

By an averaging argument, where we fix the y-part of the input to some value, there exists some $(\ell \cdot t)$-variate $\mathcal{C}_{\pm}$-circuit C' of size $n^k \leq \ell^{k(b+2)}$ such that

$$\left\langle L_\ell^{\oplus t}, C'' \right\rangle > \frac{1}{n^k}.$$

Note that for $\delta = 1/\ell^b$ and our choice of $t = \ell^b \log^2(\ell)$, we have

$$\frac{1}{n^k} > (1-2\delta)^t + \frac{1}{2\delta \cdot \ell^{k(b+2)} \cdot \ell^b}.$$

By Theorem 13, there is a $\mathsf{SUM} \circ \mathcal{C}_{\pm}$ C where $\|C\|_{\infty} \leq 1$, the top SUM has complexity $10 \cdot \ell \cdot \left(\ell^{k(b+2)} \cdot \ell^b\right)^2 \leq \ell^{O(kb)}$ and the bottom layer $\mathcal{C}_{\pm}$-circuits have size $\ell^{k(b+2)}$ such that

$$\langle L_\ell, C \rangle > 1 - 2\delta,$$

for all large enough ℓ. This contradicts Equation (1).

□

3.3 A weak equivalence theorem in the constant-error regime

In this subsection, we show a weaker equivalence theorem for the constant-error regime.

Theorem 14 (Non-black-box equivalences in the low-error regime)**.** *Let $\varepsilon \in (0, 0.5)$ and $\mathcal{C}$ be a circuit class that satisfies the following:*

- *$\mathcal{C}$ is closed under negation and projection.*
- *$\bigcup_{k\geq 1} \mathcal{C}[n^k] \subseteq \mathsf{NC}^1$.*

Then the following statements are equivalent:

1. *There is $L \in \mathsf{E}$ such that for every $k \geq 1$, L cannot be $(1/2+\varepsilon)$-approximated by $\mathcal{C}[n^k]$ under the uniform distribution.*
2. *There is $L \in \mathsf{E}$ such that, for every $k \geq 1$, L cannot be computed by a probabilistic $\mathcal{C}[n^k]$circuit with error $1/2 - \varepsilon$.*
3. *There is $L \in \mathsf{E}$ and a distribution $\mathcal{D}$ such that for every $k \geq 1$, L cannot be $(1/2 + \varepsilon)$-approximated by $\mathcal{C}[n^k]$ under $\mathcal{D}$.*
4. *There is an i.o. ε-PRG against $\mathcal{C}$ with seed length $n - 1$ and error ε.*

Proof. The equivalence of Item 2 and Item 3 follows from Yao's minimax theorem. In the following, we show Item 4 $\Rightarrow$ Item 2, Item 2 $\Rightarrow$ Item 1 and Item 1 $\Rightarrow$ Item 4, which will complete the proof.

Item 4 $\Rightarrow$ Item 2. Let $G = \{G_n\}$ be the i.o.-PRG given by Item 4. Fix a $k \geq 1$. For every $n \geq 1$, consider $G_n \colon \{0,1\}^{n-1} \to \{0,1\}^n$, and define the function $f_n \colon \{0,1\}^n \to \{0,1\}$ where $f_n(x) \coloneqq \mathbb{1}_{x\in \mathrm{Im}(G_n)}$. Suppose that for all sufficiently large n, f_n has a probabilistic $\mathcal{C}[n^k]$-circuit with error $1/2-\varepsilon$, which we denote by $\mathcal{D}_n$. Then we have for every $x \in \{0,1\}^n$ that $|\mathbf{E}_{\mathcal{D}_n}[\mathcal{D}_n(x)]-f(x)| \leq 1/2 - \varepsilon$.

Let $p = \frac{|\mathrm{Im}(G_n)|}{2^n}$. Note that $p \leq 1/2$. We have

$$\begin{aligned}
\mathop{\mathbf{E}}_{\mathcal{D}_n, x\sim\mathcal{U}_{n-1}}[\mathcal{D}_n(G_n(x))] &= \sum_{x\in\{0,1\}^{n-1}} \frac{1}{2^{n-1}} \mathop{\mathbf{E}}_{\mathcal{D}_n}[\mathcal{D}_n(G_n(x))] \\
&= \sum_{x\in\mathrm{Im}(G_n)} \frac{|G_n^{-1}(x)|}{2^{n-1}} \mathop{\mathbf{E}}_{\mathcal{D}_n}[\mathcal{D}_n(x)] \\
&= \left(\sum_{x\in\mathrm{Im}(G_n)} \frac{2\cdot|G_n^{-1}(x)|-1}{2^n} \mathop{\mathbf{E}}_{\mathcal{D}_n}[\mathcal{D}_n(x)]\right) + \left(\sum_{x\in\mathrm{Im}(G_n)} \frac{1}{2^n} \mathop{\mathbf{E}}_{\mathcal{D}_n}[\mathcal{D}_n(x)]\right) \\
&\geq (1/2+\varepsilon)\cdot\left(\sum_{x\in\mathrm{Im}(G_n)} \frac{2\cdot|G_n^{-1}(x)|-1}{2^n}\right) + p\cdot \mathop{\mathbf{E}}_{\mathcal{D},x\sim\mathcal{U}_n}[\mathcal{D}_n(x) \mid x\in\mathrm{Im}(G_n)] \\
&= \frac{(1/2+\varepsilon)}{2^n}\cdot(2^n - |\mathrm{Im}(G_n)|) + p\cdot \mathop{\mathbf{E}}_{\mathcal{D},x\sim\mathcal{U}_n}[\mathcal{D}_n(x) \mid x\in\mathrm{Im}(G_n)] \\
&= (1-p)\cdot(1/2+\varepsilon) + p\cdot \mathop{\mathbf{E}}_{\mathcal{D},x\sim\mathcal{U}_n}[\mathcal{D}_n(x) \mid x\in\mathrm{Im}(G_n)].
\end{aligned}$$

Moreover,

$$\mathop{\mathbf{E}}_{\mathcal{D}_n, x\sim\mathcal{U}_n}[\mathcal{D}_n(x)] \le (1-p)\cdot(1/2-\varepsilon) + p\cdot \mathop{\mathbf{E}}_{\mathcal{D}, x\sim\mathcal{U}_n}[\mathcal{D}_n(x) \mid x \in \mathrm{Im}(G_n)].$$

Combining these two, we have

$$\mathop{\mathbf{E}}_{\mathcal{D}_n, x\sim\mathcal{U}_{n-1}}[\mathcal{D}_n(G_n(x))] - \mathop{\mathbf{E}}_{\mathcal{D}_n, x\sim\mathcal{U}_n}[\mathcal{D}_n(x)] \ge \varepsilon.$$

Therefore, there is a circuit in the support of $\mathcal{D}_n$ that breaks the PRG G_n. Since the argument holds for every $n \ge 1$, we conclude that $\{G_n\}_n$ is *not* an i.o.-PRG, a contradiction. Therefore, for infinitely many n, f_n does not have probabilistic $\mathcal{C}[n^k]$-circuits with error $1/2-\varepsilon$.

Item 2 ⇒ Item 1. We consider two cases.

1. If DCMD cannot be $(1/2+\varepsilon)$-approximated by $\mathcal{C}[n^k]$ for every $k \ge 1$ under the uniform distribution, then we are done.
2. Otherwise, there is some $k \ge 1$ such that DCMD can be $(1/2+\varepsilon)$-approximated by $\mathcal{C}[n^k]$-circuits. By Theorem 6, CMD has a probabilistic $\mathcal{C}[n^{O(k)}]$-circuit with error $1/2-\varepsilon$. Then by Theorem 7, it implies that every function in NC^1 has a polynomial-size probabilistic $\mathcal{C}$-circuit with error $1/2-\varepsilon$. By the assumed Item 2, there is a function L in E that does not have polynomial-size $\mathcal{C}$-circuit with error up to $1/2-\varepsilon$, which implies that L is not in NC^1. Using the standard hardness amplification (Lemma 10), there is a function in E that is strongly average-case hard against NC^1, which contains polynomial-size $\mathcal{C}$-circuits.

Item 1 ⇒ Item 4. We construct the PRG as $G(x) \coloneqq x \circ f(x)$ where $\circ$ denotes concatenation. The security of the PRG can be shown by the standard connection between average-case hardness and unpredictability. □

In particular, under quite minor assumptions on the circuit class $\mathcal{C}$, it follows that even *weak* average-hardness under an *arbitrary* distribution implies similar average-case hardness under the *uniform* distribution. This complements the corresponding implication from Theorem 1, which works in a different regime of parameters and makes a somewhat stronger assumption on $\mathcal{C}$.

4 Lifting worst-case ACC^0 lower bound approaches to strong correlation bounds

In this section, we prove Theorem 2. In Section 4.1 we introduce some necessary technical ingredients for the proof. In Section 4.2 we prove the important technical lemma that $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0$-circuits can be approximated by low-degree torus polynomials. In Section 4.3 we finish the proof of Theorem 2.

4.1 Preliminaries

For three numbers a, b and $c > 0$, we write $a = b \pm c$ to mean that $b - c \le a \le b + c$.

In the following we recall two important technical facts about ACC^0-circuits, which will be crucial for the proofs in this section.

Theorem 15 (See e.g. [BHLR19, Theorem 19]). *Let $f\colon \{0,1\}^n \to \{0,1\}$ be computable by ACC^0-circuits of depth d and size $\mathrm{poly}(n)$. Then for any $e \geq 1$ there exists an integer polynomial $F(x)$ of degree $e^{O(d)} \cdot (\log n)^{O(d^2)}$ which satisfies the following: there is some $k \geq e$ such that*

$$\forall x \in \{0,1\}^n, F(x) = f(x)2^k + E(x) \pmod{2^{k+e}},$$

for some error $E(x) \leq 2^{k-e}$.

Note that invoking the result above for different circuits and the same parameter e does not guarantee the same value k. A simple way to achieve this is with the use of universal circuits.

Fact 16. *For any constants $d, m \geq 1$ and any $s \geq 1$, there is an encoding $\langle\cdot\rangle$ of n-variate circuits consisting of (unbounded fan-in) AND, OR and MOD_m gates with size at most s and depth at most d, and a universal circuit U consisting of (unbounded fan-in) AND, OR and MOD_m gates with size $s' = s^{O(1)}$ and depth $d' = O(d)$, such that for any such circuit C, we have*

- *$|\langle C\rangle| = s^c$, where $c > 0$ is a constant.*
- *for any $x \in \{0,1\}^n$, $U(\langle C\rangle, x) = C(x)$.*

4.2 $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0$ as torus polynomials

In this subsection, we prove that every $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0$-circuit of polynomial complexity can be approximated by some torus polynomial of degree $\mathsf{polylog}(n)$.

Lemma 17. *Let $d, m \geq 1$ be constants and let $\delta \in (0, 0.5)$. If a function $f\colon \{0,1\}^n \to \{0,1\}$ has a $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{AC}_d[m]$-circuit where the complexity of the top sum and the size of the $\mathsf{AC}_d[m]$-subcircuits are $\mathsf{poly}(n)$, then for every constant $b \geq 1$, f has a $(\delta/2 + 1/n^b)$-approximation torus polynomials of degree $\mathsf{polylog}(n)$.*

To prove Lemma 17, we will show two lemmas. The first says that we can "represent" a polynomial-complexity linear sum of ACC^0-circuits with *real* coefficients by a linear sum with *integer* coefficients that also has some good property. The second lemma then shows that a function with such a representation admits a pointwise approximation by a low-degree torus polynomial.

Lemma 18. *For any $\delta \in (0, 0.5)$ and any circuit class $\mathcal{C}$, let $f\colon \{0,1\}^n \to \{0,1\}$ be a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit where the complexity of the top sum and the size of the $\mathcal{C}$-subcircuits are $\mathsf{poly}(n)$. Then there are $S = \mathsf{poly}(n)$ circuits $C_1, \ldots, C_S \in \mathcal{C}[\mathsf{poly}(n)]$, S integers $\alpha_1, \ldots, \alpha_S$, and an integer T such that*

- *$\sum_{i\in[S]} |\alpha_i| \leq \mathsf{poly}(n)$,*
- *T is a power of 2 with $|T| \leq \mathsf{poly}(n)$, and*
- *for every input x*

$$\sum_{i=1}^{S} \alpha_i \cdot C_i(x) = T \cdot \left(f(x) \pm \delta'\right),$$

where $\delta' = \delta + \frac{1}{S}$.

Proof. The basic idea is to truncate the decimal fraction of the real coefficients in the original linear sum up to certain precision, and then multiply it by some power of 2.

Since f is a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit of complexity $\mathsf{poly}(n)$, there are $S = \mathsf{poly}(n)$ many circuits $C_1, C_2, \ldots, C_S \in \mathcal{C}[\mathsf{poly}(n)]$, and S real numbers $\beta_1, \ldots, \beta_S$ such that

$$\sum_{i=1}^{S} \beta_i \cdot C_i(x) = f(x) \pm \delta.$$

Also, since the coefficients are polynomially bounded, we can write, for every $i \in [S]$,

$$\beta_i = \mathsf{sign}(\beta_i) \cdot \left(\sum_{j=0}^{O(\log(n))} a_j \cdot 2^j + \sum_{k=1}^{\infty} \frac{b_k}{2^k} \right),$$

where $a_j, b_k \in \{0,1\}$. Now let

$$\beta_i' := \mathsf{sign}(\beta_i) \cdot \left(\sum_{j=0}^{O(\log(n))} a_j \cdot 2^j + \sum_{k=1}^{\lceil 2\log(S) \rceil} \frac{b_k}{2^k} \right).$$

Note that

$$|\beta_i - \beta_i'| \leq \sum_{k=\lceil 2\log(S)\rceil+1}^{\infty} \frac{1}{2^k} \leq \frac{1}{S^2}.$$

Therefore, for every x,

$$\left| \sum_{i=1}^{S} \beta_i' \cdot C_i(x) - \sum_{i=1}^{S} \beta_i \cdot C_i(x) \right| \leq \sum_{i=1}^{S} C_i(x) \cdot |\beta_i - \beta_i'| \leq \frac{1}{S}.$$

Now let

$$T := 2^{\lceil 2\log(S) \rceil}.$$

Note that T is a power of 2 and $|T| \leq \mathsf{poly}(n)$. Also, let

$$\alpha_i := T \cdot \beta_i'$$

Note that α_i is an integer and $|\alpha_i| \leq \mathsf{poly}(n)$. Finally, we have

$$\begin{aligned} \sum_{i=1}^{S} \alpha_i \cdot C_i(x) &= T \cdot \sum_{i=1}^{S} \beta_i' \cdot C_i(x) \\ &= T \cdot \left(\left(\sum_{i=1}^{S} \beta_i \cdot C_i(x) \right) \pm 1/S \right) \\ &= T \cdot \left(f(x) \pm \delta \pm 1/S \right), \end{aligned}$$

as desired. □

Lemma 19. *Let $d, m \geq 1$ be constants and let $f\colon \{0,1\}^n \to \{0,1\}$ be a function with the following property: there are $S = \mathsf{poly}(n)$ polynomial-size $\mathsf{AC}_d[m]$-circuits $C_1, \ldots, C_S$, S integers $\alpha_1, \ldots, \alpha_S$, and an integer T such that*

- $\sum_{i\in[S]} |\alpha_i| \le \mathsf{poly}(n)$,
- *T is a power of 2 with $|T| \le \mathsf{poly}(n)$, and*
- *for every input x*

$$\sum_{i=1}^{S} \alpha_i \cdot C_i(x) = T \cdot (f(x) \pm \delta)\,.$$

Then for every constant $b \ge 1$, f has a $(\delta/2 + 1/n^b)$-approximation torus polynomials of degree $\mathsf{polylog}(n)$.

Proof. Let U be the universal circuit for $\mathsf{AC}_d[m]$-circuits of size at most s from Fact 16, where s is the maximum size of $C_1, \ldots, C_S$. Applying Theorem 15 with the parameter $e = \lceil \log^2(n) \rceil$ to U, we get an integer polynomial $F(x)$ of degree $\mathsf{polylog}(n)$ and some number $k \ge e$ such that for every $\mathsf{AC}_d[m]$-circuit C of size at most s and every input $x \in \{0,1\}^n$,

$$F(\langle C\rangle, x) = U(\langle C\rangle, x) \cdot 2^k + E(\langle C\rangle, x) \pmod{2^{k+e}},$$

for some $E(\langle C\rangle, x) \le 2^{k-e}$. Let

$$p(x) := \frac{1}{2 \cdot 2^k \cdot T} \cdot \sum_{i=1}^{S} \alpha_i \cdot F(\langle C_i\rangle, x).$$

Note that p has degree $\mathsf{polylog}(n)$, since the restricted function $F(\langle C_i\rangle, \cdot)$ is a polynomial of degree $\mathsf{polylog}(n)$. We claim that $p(x)$ approximates f in the torus. To see this, note that we have

$$\begin{aligned}
q(x) &:= \sum_{i=1}^{S} \alpha_i \cdot F(\langle C_i\rangle, x) && \pmod{2^{k+e}}\\
&= \sum_{i=1}^{S} \alpha_i \cdot \Big(U(\langle C_i\rangle, x) \cdot 2^k + E(\langle C_i\rangle, x)\Big) && \pmod{2^{k+e}}\\
&= \sum_{i=1}^{S} \alpha_i \cdot \Big(C_i(x) \cdot 2^k + E(\langle C_i\rangle, x)\Big) && \pmod{2^{k+e}}\\
&= \left(2^k \cdot \sum_{i=1}^{S} \alpha_i \cdot C_i(x)\right) + \left(\sum_{i=1}^{S} \alpha_i \cdot E(\langle C_i\rangle, x)\right) && \pmod{2^{k+e}}\\
&= 2^k \cdot (T \cdot (f(x) \pm \delta)) + \left(\sum_{i=1}^{S} \alpha_i \cdot E(\langle C_i\rangle, x)\right) && \pmod{2^{k+e}}\\
&= 2^k \cdot T \cdot f(x) \pm 2^k \cdot T \cdot \delta + \left(\sum_{i=1}^{S} \alpha_i \cdot E(\langle C_i\rangle, x)\right) && \pmod{2^{k+e}}
\end{aligned}$$

Therefore, we can write

$$q(x) = 2^k \cdot T \cdot f(x) \pm 2^k \cdot T \cdot \delta + E'(x) + r(x) \cdot 2^{k+e},$$

where $|E'(x)| \leq \mathsf{poly}(n) \cdot 2^{k-e}$ and $r(x)$ is an integer. Finally, we have

$$p(x) = \frac{q(x)}{2 \cdot 2^k \cdot T} = \frac{f(x)}{2} \pm \frac{\delta}{2} + \frac{E'(x)}{2 \cdot 2^k \cdot T} + r(x) \cdot \frac{2^e}{2 \cdot T}.$$

Since T is a power of 2, and by our choice of e, the last summand $r(x) \cdot \frac{2^e}{2 \cdot T}$ is an integer. Also, since $|E'(x)| \leq \mathsf{poly}(n) \cdot 2^{k-e}$, the second last summand $\left|\frac{E'(x)}{2 \cdot 2^k \cdot T}\right| \leq \frac{\mathsf{poly}(n)}{2^e}$. As a result, we have

$$p(x) = \frac{f(x)}{2} \pm \frac{\delta}{2} \pm \frac{1}{n^{\omega(1)}} \quad (\text{over } \mathbb{T}),$$

as desired. □

Proof of Lemma 17. The lemma follows directly from Lemma 18 and Lemma 19. □

4.3 Proof of Theorem 2

Now we are ready to prove Theorem 2 (restated below).

Reminder of Theorem 2. *Consider the following statements:*

1. **Torus Polynomials:** *There is a language $L \in \mathsf{E}$ and a function $\delta(n) \geq 1/\mathsf{poly}(n)$ such that L does not have δ-approximation torus polynomials of degree $\mathsf{polylog}(n)$.*

2. **NOF Protocols:** *There is a language in E that does not admit (single-round) NOF multi-party protocols with $\mathsf{polylog}(n)$ parties of communication cost $\mathsf{polylog}(n)$.*

In each case, if the corresponding statement holds then there is a language in E that cannot be $(1/2 + 1/\mathsf{poly}(n))$-approximated under the uniform distribution by ACC^0.

Proof.

1. Torus Polynomials. By the assumption of the theorem and Lemma 17, we get that for some $\delta \geq 1/\mathsf{poly}(n)$ and for every $d, m \geq 1$, the language L cannot be computed by a $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{AC}_d[m]$-circuit where the complexity of the top sum and the size of the $\mathsf{AC}_d[m]$-subcircuits are $\mathsf{poly}(n)$. In other words, there is $\delta = 1/\mathsf{poly}(n)$ such that $\mathsf{E} \not\subseteq \widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0$. Then by Theorem 1 (from Item 2 to Item 6), there is a language in E that is strongly average-case hard against ACC^0.

2. NOF Protocols. Note that NOF multi-party protocols with $\mathsf{polylog}(n)$ parties and $\mathsf{polylog}(n)$ communication cost can simulate $\mathsf{MAJ} \circ \mathsf{ACC}^0$-circuits [BT94] (see also [Wil18b, Theorem 2.2]). This item then follows from Theorem 1 (from Item 3 to Item 6). □

5 Lower bounds against circuits with MAJ, THR, and $\widetilde{\mathsf{SUM}}$ gates

In this section, we prove Theorem 3 and Corollary 4. In Section 5.1 we introduce some important technical ingredients for our proofs in this section. In Section 5.2 we prove Item (1) of Theorem 3 and Item (1) of Corollary 4. In Section 5.3 we prove Item (2) of Theorem 3 and Items (2) and (3) of Corollary 4.

5.1 Preliminaries

For a $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$-circuit, we refer to the largest absolute value of the coefficients in the top SUM-gate as its *magnitude*, the sparsity of the top SUM-gate as its *sparsity*, and the largest size of the sub-$\mathcal{C}$-circuits as its *size*.

Error reduction for approximate linear sums. The following error reduction lemma will be useful in our proof for the Item (2) of Theorem 3.

Lemma 20. *Let $\mathcal{C}$ be a circuit class. Suppose f has a $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit where the top SUM-gate has sparsity S, magnitude T, and constant error $\delta \in (0, 0.5)$. Then for every $\varepsilon > 0$, f has an equivalent $\widetilde{\mathsf{SUM}} \circ \mathsf{AND}_{O_\delta(\log(1/\varepsilon))} \circ \mathcal{C}$-circuit with*

- *error ε,*
- *sparsity $S^{O_\delta(\log(1/\varepsilon))}$, and*
- *magnitude $S^{O_\delta(\log(1/\varepsilon))} \cdot T$.*

Proof. The idea is to reduce the error using the polynomial computing MAJ. For $0 \le \alpha \le 1$, let $\mathsf{Ber}(\alpha)$ be the distribution over $\{0,1\}$ such that $\mathbf{Pr}_{y \sim \mathsf{Ber}(\alpha)}[y = 1] = \alpha$. The following claim is a well-known fact, which follows from a standard application of Chernoff bound.

Claim 21. *There is an abosolute constant K such that the following holds. For every constant $\alpha > 0$ and every $\varepsilon > 0$, let $t = \left\lceil \frac{K}{\alpha^2} \log(1/\varepsilon) \right\rceil$. Then it holds that*

$$\mathop{\mathbf{Pr}}_{x_1,\ldots,x_t \sim \mathsf{Ber}(1/2+\alpha)}[\mathsf{MAJ}(x_1, \ldots, x_t) = 1] \ge 1 - \varepsilon$$

and

$$\mathop{\mathbf{Pr}}_{x_1,\ldots,x_t \sim \mathsf{Ber}(1/2-\alpha)}[\mathsf{MAJ}(x_1, \ldots, x_t) = 1] \le \varepsilon.$$

Now we start our proof. Let $\widetilde{f}\colon \{0,1\}^n \to \mathbb{R}$ be the corresponding linear sum of $\mathcal{C}$-circuits for f. That is,

$$\widetilde{f}(x) := \sum_{j=1}^{S} \alpha_j \cdot C_j(x),$$

where for each $j \in [S]$, C_j is a $\mathcal{C}$-circuit and $|\alpha_j| \le T$. For every $\delta > 0$, let $c_\delta > 0$ be a large enough constant to be specified later. let $t = c_\delta \cdot \log(1/\varepsilon) \le O_\delta(\log 1/\varepsilon)$. Also let $p\colon \mathbb{R}^t \to \mathbb{R}$ be the multi-linear polynomial computing MAJ on t input bits. Namely, for each $x \in \{0,1\}^t$ it holds that $p(x) = \mathsf{MAJ}(x)$. Note that p has degree at most t. Define

$$q(x) := \frac{\widetilde{f}(x) + \delta}{1 + 2\delta}.$$

Note that for every $x \in \{0,1\}^n$, we have $0 \le q(x) \le 1$.

Since p is multi-linear, for every input $x \in \{0,1\}^n$, we have

$$\begin{aligned} p(q(x), q(x) \ldots, q(x)) &= \mathop{\mathbf{E}}_{y_1,y_2,\ldots,y_t \sim \mathsf{Ber}(q(x))}[p(y_1, y_2, \ldots, y_t)] \\ &= \mathop{\mathbf{Pr}}_{y_1,y_2,\ldots,y_t \sim \mathsf{Ber}(q(x))}[\mathsf{MAJ}(y_1, y_2, \ldots, y_t) = 1]. \end{aligned} \tag{2}$$

We now show that the function

$$P(x) := p(q(x), q(x) \ldots, q(x))$$

ε-approximates f. If $f(x) = 0$, then $\widetilde{f}(x) \in [-\delta, \delta]$, which implies $0 \leq q(x) < \frac{2\delta}{1+2\delta} < 1/2$. Otherwise, we have $f(x) = 0$ and $\widetilde{f}(x) \in [1-\delta, 1+\delta]$, which implies that $\frac{1}{2} < \frac{1}{1+2\delta} < q(x) \leq 1$. Now, by Claim 21 and (2), for large enough constant c_δ and $t = c_\delta \log(1/\varepsilon)$, it holds that $P(x) \in [0, \varepsilon]$ for $q(x) < \frac{1}{2}$ and $P(x) \in [1-\varepsilon, 1]$ for $q(x) > 1/2$.

Since p is a multi-linear polynomial, we can write $P(x)$ as $P(x) = \sum_{i=0}^{t} \beta_i \cdot q(x)^i$ for proper coefficients $(\beta_i)_{i=0}^{t}$. Being an affine transformation of $\widetilde{f}$, q can also be computed by a sum of $\mathcal{C}$-circuits of sparsity S. Then, for each $i \in \{0, \ldots, t\}$ we can compute $q(x)^i$ by a sum of $\mathcal{C}$-circuits of sparsity at most S^i. Hence, P can be written as a linear sum of $\sum_{i=0}^{t} S^i = O(S^t)$ $\mathsf{AND}_t \circ \mathcal{C}$-circuits. This justifies the sparsity requirement in the conclusion of the lemma. Verifying the magnitude requirement is straightforward. This completes the proof. □

Efficient construction of probabilistic proofs systems (PCPs). We will make use of the following PCP construction from [Vio20].

Lemma 22 ([Vio20, Lemma 11])**.** *Let M be an algorithm running in time $T = T(n) \geq n$ on inputs of the form (x, y) where $|x| = n$. Given $x \in \{0,1\}^n$ one can output in time $\mathsf{poly}(n, \log T)$ a collection of $\mathsf{poly}(r)$ circuits $q_j^i \colon \{0,1\}^r \to \{0,1\}^r$ for $i \in [\mathsf{poly}(r)]$ and $j \in [3]$, and $R_i \colon \{0,1\}^3 \to \{0,1\}$ such that:*

- ***Proof length.*** *$2^r \leq T \cdot \mathsf{polylog}(T)$.*
- ***Completeness.*** *If there is y such taht $M(x, y)$ accepts then there is a map $\pi\{0,1\}^r \to \{0,1\}$ such that for any $z \in \{0,1\}^r$ and $i \in [\mathsf{poly}(r)]$ we have $R_i(\pi(q_1^i(z)), \pi(q_2^i(z)), \pi(q_3^i(z))) = 1$.*
- ***Soundness.*** *If no y causes $M(x, y)$ to accept, then for every map $\pi \colon \{0,1\}^r \to \{0,1\}$, at most $1 - 1/r^{O(1)}$ fraction of the pairs $(z, i) \in \{0,1\}^r \times [\mathsf{poly}(r)]$ have that $R_i(\pi(q_1^i(z)), \pi(q_2^i(z)), \pi(q_3^i(z))) = 1$.*
- ***Complexity.*** *Each q_j^i is a projection (a.k.a. 1-local), i.e., each output bit of q_j^i is one input bit, the negation of an input bit, or a constant; each R_i is an OR of three literals.*

Tests on approximate linear sums. We also need the following "close-to-Boolean" test, adapted from [CW19].

We start with some definitions. For a function $f \colon \{0,1\}^n \to \mathbb{R}$, we define the Boolean function associated with f as $\mathsf{bin}(f) = \mathbb{1}_{f(x) \geq 1/2}$. For any function $f \colon \{0,1\}^n \to \mathbb{R}$ and a real $p \geq 1$, we define its ℓ_p-norm as

$$\|f\|_p := \left(\mathop{\mathbf{E}}_{x \sim \mathcal{U}_n} |f(x)|^p \right)^{1/p}.$$

We also define the ℓ_∞-norm as

$$\|f\|_\infty := \max_{x \in \{0,1\}^n} |f(x)|.$$

Lemma 23 (Following [CW19, Lemma 33])**.** *For any integer $S > 0$, let*

$$f = \sum_{i \in [S]} \alpha_i \cdot C_i,$$

where for every $i \in [S]$, $\alpha_i \in \mathbb{R}$ and C_i is a $\mathcal{C}$-circuit. Suppose the #SAT problem on $\mathsf{AND}_8 \circ \mathcal{C}$-circuits of n-bit inputs can be computed in time $T(n)$. Then for any $\varepsilon \in (0, 0.01)$, there is an algorithm A running in time $O\left(T(n) \cdot S^8\right)$ such that:

- *if $\|f - \mathsf{bin}(f)\|_\infty \leq \varepsilon$, then A accepts;*
- *if $\|f - \mathsf{bin}(f)\|_4 \geq 3 \cdot \varepsilon$, then A rejects;*
- *otherwise, A can output anything.*

Proof. Let $P \colon \mathbb{R} \to \mathbb{R}$ be a degree-8 polynomial defined as

$$P(z) \coloneqq z^4 \cdot (1 - z)^4.$$

It is easy to verify that for any $z \in \mathbb{R}$, letting $d_{\mathsf{bin}}(z) \coloneqq |z - \mathsf{bin}(z)|$, we have

- $P(z) \leq d_{\mathsf{bin}}(z)^4 \cdot (1 + d_{\mathsf{bin}}(z))^4$ and
- $P(z) \geq d_{\mathsf{bin}}(z)^4 \cdot 2^{-4}$.

On the one hand, if $\|f - \mathsf{bin}(f)\|_\infty \leq \varepsilon$, then $d_{\mathsf{bin}}(f(x)) \leq \varepsilon$ for every x, which implies

$$\mathop{\mathbf{E}}_x[P(f(x))] \leq \varepsilon^4 \cdot (1 + \varepsilon)^4 \leq \varepsilon^4 \cdot 1.01^4.$$

On the other hand, if $\|f - \mathsf{bin}(f)\|_4 \geq 3 \cdot \varepsilon$, then we have

$$\mathop{\mathbf{E}}_x[P(f(x))] \geq 2^{-4} \cdot \mathop{\mathbf{E}}_x\left[d_{\mathsf{bin}}(f(x))^4\right] = 2^{-4} \cdot \|f - \mathsf{bin}(f)\|_4^4 \geq (3/2)^4 \cdot \varepsilon^4.$$

Therefore, to distinguish the two cases, it suffices to compute

$$\mathop{\mathbf{E}}_x[P(f(x))],$$

which can be done by making $O\left(S^8\right)$ calls to the #SAT algorithm for $\mathsf{AND}_8 \circ \mathcal{C}$-circuits. □

Finally, the following inequality will also be useful for us.

Lemma 24 (Cauchy-Schwarz)**.** *For any functions $f, g \colon \{0,1\}^n \to \mathbb{R}$ and integer $t \geq 1$, we have*

$$\|f \cdot g\|_t \leq \|f\|_{2t} \cdot \|g\|_{2t}.$$

Proof. We have

$$\begin{aligned}
\|f \cdot g\|_t &= \mathop{\mathbf{E}}_x\left[|f(x) \cdot g(x)|^t\right]^{1/t} \\
&\leq \mathop{\mathbf{E}}_x\left[|f(x)|^t \cdot |g(x)|^t\right]^{1/t} \\
&= \left\langle |f|^t, |g|^t \right\rangle^{1/t} \\
&\leq \left\||f|^t\right\|_2^{1/t} \cdot \left\||g|^t\right\|_2^{1/t} \\
&= \mathop{\mathbf{E}}_x\left[|f(x)|^{2t}\right]^{1/(2t)} \cdot \mathop{\mathbf{E}}_x\left[|g(x)|^{2t}\right]^{1/(2t)} \\
&= \|f\|_{2t} \cdot \|g\|_{2t},
\end{aligned}$$

as desired. □

5.2 $\mathsf{LTF}^{\mathbf{quasipoly}(n)} \circ \mathcal{C}$ lower bounds

In this section we prove Item (1) of Theorem 3 (restated below). Recall that a circuit class $\mathcal{C}$ is *nice* if $\mathcal{C}$ is closed under negation, (bottom) projections, and a top AND gate of unbounded fan-in, and in addition $\mathcal{C}$-circuits of size s admit general circuits of depth $O(\log s)$.

Reminder of Item (1) of Theorem 3. *Let $\mathcal{C}$ be a nice circuit class. Suppose there is a constant $\varepsilon > 0$ such that, given a $\mathcal{C}$-circuit of size 2^{n^ε} over n input variables, its number of satisfying assignments can be computed in time 2^{n-n^ε}. Then for every constant $k > 0$, NQP does not have $\mathsf{LTF}^{2^{\log^k n}} \circ \mathcal{C}$-circuits of size $2^{\log^k n}$.*

First, we state the following useful lemma.

Lemma 25. *Let $\mathcal{C}$ be a circuit class that is closed under negation and projection. Let $s\colon \mathbb{N} \to \mathbb{N}$ be a non-decreasing function computable in polynomial time. If DCMD can be computed by an ensemble of $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits $\{C_n\}$ of size $s(n)$, then for every $\delta \in (0, 0.5)$, CMD can be computed by an ensemble of $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits of size $\mathsf{poly}(n, s(O(n^2)), 1/\delta)$.*

Before proving Lemma 25, we show how it implies Item (1) of Theorem 3.

Proof Sketch of Item (1) of Theorem 3. Let $s(n) := 2^{\log^k n}$. Depending on the complexity of DCMD, there are two cases.

1. If DCMD cannot be computed by $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits of size at most $2^{\log^k n}$, then we are done since DCMD is itself in NQP.

2. Otherwise, DCMD can be computed by $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits of size at most $2^{\log^k n}$. By Lemma 25 we know that CMD can be computed by $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$-circuits of size $\mathsf{poly}(n, s(O(n^2)))$. By Theorem 7, NC^1 collapses to $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$-circuits of quasi-polynomial size.

   We choose d to be a large enough constant such that $s(n)$-size $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits can be simulated by (general) circuits of depth at most $\log^d n$.[13] Since NC^1 collapses to $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$ of quasi-polynomial size, we can then choose d' such that every circuit of depth at most $\log^d n$ can be simulated by a $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$-circuit of size at most $2^{\log^{d'} n}$.

   Assuming the 2^{n-n^ε}-time CAPP algorithm for 2^{n^ε} size $\mathcal{C}$-circuits, it follows from [CW19] and [CR20] that NQP cannot be computed by $\widetilde{\mathsf{SUM}}_{1/3} \circ \mathcal{C}$-circuits of size $2^{\log^{d'} n}$, which in turn implies that NQP cannot be computed by $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits of size at most $2^{\log^k n}$. This completes the proof.

□

Now we prove Lemma 25.

[13] Using similar reasoning as in [CW19, Propositon 40], we can assume the top LTF gate has coefficients being rationals with numerator and denominator bounded by $2^{\mathsf{poly}(s)}$, where s is the sparsity of the top gate. Then, given that $\mathcal{C}$ is a nice circuit class, $\mathcal{C}$-circuits of size s can be simulated in depth $O(\log s)$, and it is clear that we can simulate $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$ by low-depth circuits.

Proof of Lemma 25. We use the randomized reduction from DCMD to CMD. Suppose that there is a $\mathsf{LTF}^s \circ \mathcal{C}$-circuit computing DCMD. By Theorem 6, for every $n \geq 1$, there is an $m = m(n) \leq O(n^2)$ and a random reduction from CMD_n to DCMD_m. That is, there is a function $R\colon \{0,1\}^n \times \{0,1\}^{O(n^2)} \to \{0,1\}^m$ such that the following hold.

1. For every $r \in \{0,1\}^{O(n^2)}$, it holds that $\mathsf{CMD}(x) = r_0 \oplus \mathsf{DCMD}(R(x,r))$ where r_0 denotes the first bit of r.
2. For every $y \in \{0,1\}^m$, $\mathbf{Pr}_{r \in \{0,1\}^{O(n^2)}}[R(x,r) = y] = 2^{-m}$.
3. R computes a projection of input bits.

Let $C = \sum_{i=1}^t \alpha_i C_i$ be the assumed $s(m)$-size $\mathsf{LTF}^{s(m)} \circ \mathcal{C}$-circuit for DCMD_m. We let

$$V := \mathop{\mathbf{E}}_{x \sim \{0,1\}^m}[|C(x)|].$$

It follows from the definition of $\mathsf{LTF}^{s(n)} \circ \mathcal{C}$-circuits that $V \geq 1/s$.

Now we construct a *probablistic* $\mathsf{SUM} \circ \mathcal{C}$-circuit $\mathcal{D}$ to compute CMD_n as follows.

- On an input $x \in \{0,1\}^n$, $\mathcal{D}$ samples a uniform random string $r \in \{0,1\}^{O(n^2)}$, and outputs $C(R(x,r)) \cdot (-1)^{r_0}$.

Let $P_{\mathcal{D}}(x)$ denote the expectation of output of $\mathcal{D}$ given $x \in \{0,1\}^n$ as input. By the random reduction property we have that $P_{\mathcal{D}}(x) = V \cdot (-1)^{\mathsf{CMD}_n(x)}$. For brevity, we use $f\colon \{0,1\}^n \to \{0,1\}$ to denote CMD_n in the following.

Since C is a $\mathsf{LTF}^{s(m)} \circ \mathcal{C}$-circuit, the output of $\mathcal{D}$ is always bounded by $s(m)$. Hence, we choose $\ell = \Theta\left(s(m)^4 \cdot n \cdot \delta^{-4}\right)$ and sample ℓ circuits from $\mathcal{D}$, denoted by $D_1, \ldots, D_\ell$. By a Chernoff bound and union bound, we can fix a choice of $D_1, \ldots, D_\ell$ such that for every $x \in \{0,1\}^n$, it holds that

$$\left| \frac{1}{\ell} \sum_{i=1}^{\ell} D_i - V \cdot (-1)^{f(x)} \right| \leq \frac{\delta}{5s(m)}. \tag{3}$$

Define $D(x) := \frac{1}{\ell} \sum_{i=1}^{\ell} D_i(x)$. Note that D is a $\mathsf{SUM} \circ \mathcal{C}$-circuit. Finally, we define

$$D'(x) := \frac{1}{2} + \frac{D(x)}{2\left(V + \frac{\delta}{s}\right)}$$

which is just an affine transformation of D. Then, it follows that for every $x \in \{0,1\}^n$, we have:

$$\begin{aligned}
|D'(x) - f(x)| &\leq \frac{1}{2(V+\delta/s)} \left| D(x) + (V + \delta/s) - 2 \cdot f(x)(V + \delta/s) \right| \\
&\leq \frac{1}{2(V+\delta/s)} \left(\left| V \cdot (-1)^{f(x)} + (V + \delta/s) - 2 \cdot f(x)(V + \delta/s) \right| + \frac{\delta}{5s} \right) \\
&\leq \frac{2\delta/s}{2(V+\delta/s)} \\
&\leq \delta. && (V \geq 1/s)
\end{aligned}$$

□

Finally, combing the 2^{n-n^ε} time #SAT algorithm for 2^{n^ε}-size $\mathsf{ACC}^0 \circ \mathsf{THR}$-circuits from [Wil18b], Item (1) of Corollary 4 (restated below) follows immediately.

Reminder of Item (1) of Corollary 4. *For every constant $k > 0$, NQP does not admit $\mathsf{LTF}^{2^{\log^k n}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits of size $2^{\log^k n}$.*

5.3 Average-case lower bounds against $\widetilde{\mathsf{SUM}} \circ \mathsf{ACC}^0 \circ \mathsf{THR}$

Next we prove Item (2) of Theorem 3 (restated below).

Reminder of Item (2) of Theorem 3. *Let $\mathcal{C}$ be a nice circuit class. Suppose there is a constant $\varepsilon > 0$ such that, given a $\mathcal{C}$-circuit of size 2^{n^ε} over n input variables, its number of satisfying assignments can be computed in time 2^{n-n^ε}. Then for every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be $(1/2 + 2^{-\log^k n})$-approximated by $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits where both the sparsity of the top SUM-gate and the size of the bottom layer $\mathcal{C}$-circuits are at most $2^{\log^k n}$.*

Although Items (1) and (2) of Theorem 3 seem incomparable, we observe that both of them imply a quasi-polynomial size $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}} \circ \mathcal{C}$ lower bound, as sketched in the introduction of the paper.

Again, applying the 2^{n-n^ε}-time #SAT algorithm for size-2^{n^ε} $\mathsf{ACC}^0 \circ \mathsf{LTF}$-circuits, Item (2) of Corollary 4 follows immediately. In turn, Item (3) of Corollary 4 follows from Item (2) and a standard application of the Discriminator Lemma.

Reminder of Items (2) and (3) of Corollary 4. *The following results hold:*

2 *For every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be $(1/2 + 2^{-\log^k n})$-approximated by $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits where the top SUM-gate has sparsity $2^{\log^k n}$ and all $\mathsf{ACC}^0 \circ \mathsf{THR}$-subcircuits have size $2^{\log^k n}$.*

3 *For every choice of constants $k > 0$ and $\delta \in (0, 0.5)$, NQP cannot be computed by $\mathsf{MAJ} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-circuits where the top MAJ gate has fan-in $2^{\log^k n}$ and all $\widetilde{\mathsf{SUM}}_\delta \circ \mathsf{ACC}^0 \circ \mathsf{THR}$-subcircuits have size and sparsity $2^{\log^k n}$.*

The proof of Item (2) of Theorem 3 follows the framework of [CR20]. In particular, we prove the following lemma in substitution of [CR20, Theorem 5.1] in the original proof.

Lemma 26. *Let $\mathcal{C}$ be a nice circuit class. There is a universal constant $\delta_{\mathsf{CMD}} > 0$ such that, if the followings hold:*

1. *there is a constant $\delta \in (0, 0.5)$ such that CMD can be computed by a family of $\widetilde{\mathsf{SUM}}_{\delta_{\mathsf{CMD}}} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits, where the top $\widetilde{\mathsf{SUM}}$ has sparsity and magnitude at most $2^{\log^{O(1)} n}$, the middle $\widetilde{\mathsf{SUM}}_\delta$ gates have sparsity at most $2^{\log^{O(1)} n}$, and the bottom $\mathcal{C}$-circuits have size at most $2^{\log^{O(1)} n}$, and*
2. *there is a constant $\varepsilon > 0$ such that given a $\mathcal{C}$-circuit on n-bit inputs of size at most 2^{n^ε}, the number of satisfying assignments to the circuit can be computed in time 2^{n-n^ε},*

then NE *can certify* $n^{\Omega(1)}$ *depth hardness. Namely, there is an* $O(2^n)$ *time algorithm* $\mathcal{A}(x,y)$ *such that, for infinitely many* n*: (1)* $\mathcal{A}(1^n, tt(f)) = 1$ *for a function* $f\colon \{0,1\}^n \to \{0,1\}$ *implies that* f *does not admit* $n^{o(1)}$*-depth circuits and (2)* $\mathcal{A}(1^n, tt(f)) = 1$ *for some* n*-bit function* f*.*

Given Lemma 26, Item (2) of Theorem 3 can be proved in the same way as that of [CR20].

Proof Sketch of Item (2) of Theorem 3. For the sake of contradiction, if there exists some k such that NQP can be $(1/2 + 2^{-\log^k n})$-approximated by $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$ of sparsity and $\mathcal{C}$-circuit size at most $2^{\log^k n}$, then by the property of DCMD/CMD (Theorem 6 and Theorem 7), the first item in the assumption of Lemma 26 (see [CR20, Section 3]) holds. Now, with the non-trivial #SAT algorithm for $\mathcal{C}$-circuits, we can certify $n^{\Omega(1)}$ depth hardness in NE. Such a hardness certifier can be used to construct functions in NQP that cannot be $(1/2 + 2^{-\log^k n})$-approximated by (general) circuits of depth at most $\log^{O(k)} n$ (see [CR20, Section 4]). This completes the proof, since it is possible to simulate $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuits as in the statement of the theorem using circuits of depth $\log^{O(k)} n$ by upper bounding the bit complexity of their coefficients (see [CW19, Propositon 40]). □

The rest of this section is devoted to prove lemma 26.

5.3.1 Certifying low-depth circuits: Proof of Lemma 26

Given Lemma 20, the proof of Lemma 26 proceeds in the same way as that of [CR20, Theorem 5.1], using the observation that we can now derive from Item 1 of the assumption in Lemma 26 via error reduction and a collapse that CMD has a $\widetilde{\mathsf{SUM}}_{2\delta_{\mathsf{CMD}}} \circ \mathcal{C}$-circuit of *sparsity* (not *complexity*) at most $2^{\log^{O(1)} n}$. We highlight that the original proof combined the PCP of [BV14] with PCPs of proximity (PCPP) for technical reasons, while we will present a proof that only uses the PCP of [BV14] (more precisely, a slight variant presented in [Vio20]). By combining this PCP with Lemma 20, we can prove Lemma 26 without PCPP.

We are now ready to show Lemma 26.

Proof of Lemma 26. Let $\delta_{\mathsf{CMD}} < 1/8$. We have the following observation.

Claim 27. *Let* $c \geq 1$ *be a constant. Suppose Item 1 in the statement of Lemma 26 holds. Then* CMD *has a* $\widetilde{\mathsf{SUM}}_{1/n^c} \circ \mathcal{C}$*-circuit with sparsity and* $\mathcal{C}$*-circuit size* $2^{\log^{O(1)} n}$*.*

Proof of Claim 27. By the first assumption of Lemma 26, CMD has a $\widetilde{\mathsf{SUM}}_{\delta_{\mathsf{CMD}}} \circ \widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-circuit. Here the top $\widetilde{\mathsf{SUM}}$ gate has sparsity $S_1 = 2^{\log^{O(1)} n}$ and magnitude $T_1 = 2^{\log^{O(1)} n}$, the middle $\widetilde{\mathsf{SUM}}_\delta$ gates have sparsity at most $2^{\log^{O(1)} n}$ and error $\delta \in (0, 0.5)$, and the bottom $\mathcal{C}$-circuits have size at most $2^{\log^{O(1)} n}$. We first use Lemma 20 to reduce the error of each $\widetilde{\mathsf{SUM}}_\delta \circ \mathcal{C}$-subcircuit from δ to

$$\frac{1}{8 \cdot S_1 \cdot T_1}.$$

Note that this step only expands the $\mathcal{C}$-circuit size, sparsity and magnitude of the $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$-circuits by a quasi-polynomial factor. Then we replace each of these subcircuits with their corresponding linear sums and merge them with the top sum. After this we get a $\widetilde{\mathsf{SUM}}_{1/4} \circ \mathcal{C}$-circuit computing CMD with sparsity and $\mathcal{C}$-circuit size at most $2^{\log^{O(1)} n}$. Finally, we apply Lemma 20 again to reduce the error of the $\widetilde{\mathsf{SUM}} \circ \mathcal{C}$ to $1/n^c$. This step only incurs a quasi-polynomial blow-up in the circuit sparsity, magnitude and size. This completes the proof. □

Let L be a unary language in $\mathsf{NTIME}[2^n]\backslash\mathsf{NTIME}[o(2^n)]$. Consider the PCP of Lemma 22 for L. For $z \in \{0,1\}^r$ where $r = r(n) = n + O(\log n)$, and $i \in [\mathsf{poly}(r)]$, denote the queries $Q_i(z)$ by $\left(q_1^i(z), q_2^i(z), q_3^i(z)\right) \in \{0,1\}^r \times \{0,1\}^r \times \{0,1\}^r$, where q_1^i, q_2^i, q_3^i are all projection functions. By Lemma 22, we have:

- If $x \in L$, then there exists a proof $\pi\colon \{0,1\}^r \to \{0,1\}$, such that

$$\mathop{\mathbf{E}}_{z\sim\{0,1\}^r, i\sim[\mathsf{poly}(n)]}\left[R_i\left(\pi(q_1^i(z)), \pi(q_2^i(z)), \pi(q_3^i(z))\right)\right] = 1.$$

- If $x \notin L$, then for every proof $\pi\colon \{0,1\}^r \to \{0,1\}$, we have

$$\mathop{\mathbf{E}}_{z\sim\{0,1\}^r, i\sim[\mathsf{poly}(n)]}\left[R_i\left(\pi(q_1^i(z)), \pi(q_2^i(z)), \pi(q_3^i(z))\right)\right] \le 1 - 1/n^b$$

  for some absolute constant $b \ge 1$.

Note that in the above, the sets of query functions and predicates depend on x. We define a PCP verifier $V\colon \{0,1\}^n \times \{0,1\}^{2^r} \to \{0,1\}$ as follows:

$$V(x,y) = 1 \iff \mathop{\mathbf{E}}_{z\sim\{0,1\}^r, i\sim[\mathsf{poly}(n)]}\left[R_i\left(\pi_y(q_1^i(z)), \pi_y(q_2^i(z)), \pi_y(q_3^i(z))\right)\right] = 1,$$

where $\pi_y\colon \{0,1\}^r \to \{0,1\}$ is the function whose truth table is y. It is clear that V can be computed in time $2^n \cdot \mathsf{poly}(n)$.

Next, we show that V certifies $r^{\Omega(1)}$-depth hardness. Namely, for infinitely many n, we have $1^n \in L$, but for any y such that $V(1^n, y) = 1$, π_y cannot be computed by circuits of depth $n^{o(1)}$.

For the sake of contradiction, suppose for every $1^n \in L$, V on input 1^n has a witness that can be computed by circuit of depth $h = h(r) = r^{o(1)}$. We will design an algorithm to decide L in time $\mathsf{NTIME}[o(2^n)]$, which contradicts that $L \notin \mathsf{NTIME}[o(2^n)]$. The assumption above implies that, for every large enough n, if $1^n \in L$, there exists some proof oracle π^*, which has a depth-h circuit, such that

$$\mathop{\mathbf{E}}_{z\sim\{0,1\}^r, i\sim[\mathsf{poly}(n)]}\left[R_i\left(\pi^*(q_1^i(z)), \pi^*(q_2^i(z)), \pi^*(q_3^i(z))\right)\right] = 1. \tag{4}$$

Since CMD is $\oplus\mathsf{L}$-complete under projection (Theorem 7) and π^* has circuit of depth h, there is a projection function $R\colon \{0,1\}^r \to \{0,1\}^{2^{O(h)}}$ such that for every $u \in \{0,1\}^r$, it holds $\pi^*(u) = \mathsf{CMD}(R(u))$. Therefore, by Claim 27, π^* has a $\widetilde{\mathsf{SUM}}_\varepsilon \circ \mathcal{C}$-circuit with sparsity at most $S = 2^{h^{O(1)}}$, size at most $m = 2^{h^{O(1)}}$, and error $\varepsilon = 1/r^c$, where $c > 0$ is a sufficiently large constant to be chosen later.

Now we describe an algorithm $\mathcal{A}$ to decide L in $o(2^n)$ time. On an input 1^n, $\mathcal{A}$ first guesses a linear sum of at most S $\mathcal{C}$-circuits, where each $\mathcal{C}$-subcircuit has size at most m. Let's call it $\widetilde{\pi}$. $\mathcal{A}$ then runs the "close-to-Boolean" test in Lemma 23 with parameter ε on $\widetilde{\pi} \circ q_j^i$ for *every* $i \in [\mathsf{poly}(r)]$ and $j \in [3]$. It rejects immediately if any of these tests rejects. Note that this can be done using the assumed non-trivial $\#\mathsf{SAT}$ algorithm for $\mathcal{C}$-circuits. If $1^n \in L$, then the algorithm finds a linear sum of $\mathcal{C}$-circuits $\widetilde{\pi}$ that is pointwise ε-close to the "low-depth" proof oracle π^*. In this case, we have for every $i \in [\mathsf{poly}(r)]$ and $j \in [3]$ that:

- $\pi^* \circ q_j^i = \mathsf{bin}(\widetilde{\pi} \circ q_j^i)$

- $\left\|\mathsf{bin}(\widetilde{\pi} \circ q_j^i) - \widetilde{\pi} \circ q_j^i\right\|_\infty = \left\|\pi^* \circ q_j^i - \widetilde{\pi} \circ q_j^i\right\|_\infty \leq \varepsilon,$

so such a "good" guess will never be rejected by the above test. Moreover, for all the guesses that are not rejected by the test, we have, for every $i \in [\mathsf{poly}(r)]$ and $j \in [3]$,

$$\left\|\mathsf{bin}(\widetilde{\pi} \circ q_j^i) - \widetilde{\pi} \circ q_j^i\right\|_4 \leq 3 \cdot \varepsilon.$$

Note that for every $i \in [\mathsf{poly}(r)]$, the predicate function R_i can be written as a multi-linear polynomial on its inputs, i.e.,

$$R_i(u_1, u_2, u_3) = \sum_{T \subseteq [3]} a_T \cdot \prod_{j \in T} u_j,$$

where for every $T \subseteq [3]$ we have $a_T \in [-5, 5]$, since R_i maps $\{0,1\}^3$ to $\{0,1\}$. Next, we show the following claim.

Claim 28. *We have*

$$\left|\mathop{\mathbf{E}}_{z,i}\left[R_i\left(\widetilde{\pi}(q_1^i(z)), \widetilde{\pi}(q_2^i(z)), \widetilde{\pi}(q_3^i(z))\right)\right] - \mathop{\mathbf{E}}_{z,i}\left[R_i\left(\mathsf{bin}(\widetilde{\pi}(q_1^i(z))), \mathsf{bin}(\widetilde{\pi}(q_2^i(z))), \mathsf{bin}(\widetilde{\pi}(q_3^i(z)))\right)\right]\right| \leq O(\varepsilon).$$

Proof. It suffices to show that for every $i \in [\mathsf{poly}(r)]$ and $T \subseteq [3]$, the quantity

$$\left|\mathop{\mathbf{E}}_{z}\left[\prod_{j \in T} \widetilde{\pi}(q_j^i(z))\right] - \mathop{\mathbf{E}}_{z}\left[\prod_{j \in T} \mathsf{bin}(\widetilde{\pi}(q_j^i(z)))\right]\right| \tag{5}$$

is at most $O(\varepsilon)$. Let us fix $i \in [\mathsf{poly}(r)]$ and denote $\widetilde{\pi}(q_j^i(z))$ by $g_j(z)$ for each of $j \in [3]$. We upper bound (5) by cases.

$|T| = 1$. Assume without loss of generality that $T = \{1\}$, we have

$$(5) = \left|\mathop{\mathbf{E}}_{z}\left[g_1(z) - \mathsf{bin}(g_1(z))\right]\right| \leq \|g_1 - \mathsf{bin}(g_1)\|_1 \leq \|g_1 - \mathsf{bin}(g_1)\|_4 \leq 3 \cdot \varepsilon.$$

$|T| = 2$. Assume without loss of generality that $T = \{1, 2\}$. Using Lemma 24, we have

$$\begin{aligned}
\|g_1 g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_2 &\leq \|g_1 g_2 - g_1 \mathsf{bin}(g_2)\|_2 + \|g_1 \mathsf{bin}(g_2) - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_2 \\
&= \|g_1 \cdot (g_2 - \mathsf{bin}(g_2))\|_2 + \|(g_1 - \mathsf{bin}(g_1)) \cdot \mathsf{bin}(g_2)\|_2 \\
&\leq \|g_1\|_4 \cdot \|g_2 - \mathsf{bin}(g_2)\|_4 + \|g_1 - \mathsf{bin}(g_1)\|_4 \cdot \|\mathsf{bin}(g_2)\|_4 \\
&\leq (1 + 3\varepsilon) \cdot 3\varepsilon + 3\varepsilon \cdot 1 \\
&\leq 3\varepsilon(2 + 3\varepsilon),
\end{aligned}$$

where the second last inequality uses that $\|g_1\|_4 \leq \|g_1 - \mathsf{bin}(g_1)\|_4 + \|\mathsf{bin}(g_1)\|_4 \leq 1 + 3\varepsilon$. Then

$$(5) \leq \|g_1 g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_1 \leq \|g_1 g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_2 \leq 3\varepsilon(2 + 3\varepsilon).$$

$|T| = 3$. Again, using Lemma 24 , we have

$$\begin{aligned}
(5) &\leq \|g_1g_2g_3 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\mathsf{bin}(g_3)\|_1 \\
&\leq \|g_1g_2g_3 - g_1g_2\mathsf{bin}(g_3)\|_1 + \|g_1g_2\mathsf{bin}(g_3) - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\mathsf{bin}(g_3)\|_1 \\
&= \|g_1g_2 \cdot (g_3 - \mathsf{bin}(g_3))\|_1 + \|(g_1g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)) \cdot \mathsf{bin}(g_3)\|_1 \\
&\leq \|g_1g_2\|_2 \cdot \|g_3 - \mathsf{bin}(g_3)\|_2 + \|g_1g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_2 \cdot \|\mathsf{bin}(g_3)\|_2 \\
&\leq \|g_1\|_4 \cdot \|g_2\|_4 \cdot \|g_3 - \mathsf{bin}(g_3)\|_2 + \|g_1g_2 - \mathsf{bin}(g_1)\mathsf{bin}(g_2)\|_2 \cdot \|\mathsf{bin}(g_3)\|_2 \\
&\leq (1+3\varepsilon)^2 \cdot 3\varepsilon + 3\varepsilon(2+3\varepsilon) \cdot 1 \\
&= O(\varepsilon).
\end{aligned}$$

This completes the proof of Claim 28. □

On the one hand, if $1^n \in L$, then there exist a good guess $\widetilde{\pi}$ such that $\mathsf{bin}(\widetilde{\pi}) = \pi^*$ and therefore

$$\mathop{\mathbf{E}}_{z,i}\left[R_i\left(\mathsf{bin}(\widetilde{\pi}(q_1^i(z))), \mathsf{bin}(\widetilde{\pi}(q_2^i(z))), \mathsf{bin}(\widetilde{\pi}(q_3^i(z)))\right)\right] = 1,$$

which, by Claim 28, implies that

$$\mathop{\mathbf{E}}_{z,i}\left[R_i\left(\widetilde{\pi}(q_1^i(z)), \widetilde{\pi}(q_2(z)), \widetilde{\pi}(q_3(z))\right)\right] \geq 1 - O(\varepsilon) \geq 1 - O\left(1/n^c\right) \geq 1 - 1/\left(2n^b\right), \tag{6}$$

by choosing c to be a large enough constant. On the other hand, if $1^n \notin L$, then for any guess $\widetilde{\pi}$, we have

$$\mathop{\mathbf{E}}_{z,i}\left[R_i\left(\mathsf{bin}(\widetilde{\pi}(q_1^i(z))), \mathsf{bin}(\widetilde{\pi}(q_2^i(z))), \mathsf{bin}(\widetilde{\pi}(q_3^i(z)))\right)\right] \leq 1 - 1/n^b,$$

which gives

$$\mathop{\mathbf{E}}_{z,i}\left[R_i\left(\widetilde{\pi}(q_1^i(z)), \widetilde{\pi}(q_2(z)), \widetilde{\pi}(q_3(z))\right)\right] \leq 1 - 1/n^b + O(\varepsilon) < 1 - 1/\left(2n^b\right).$$

Given a #SAT algorithm for $\mathcal{C}$-circuit of size $2^{r^{\Omega(1)}}$ with running time $2^{r-r^{\Omega(1)}}$, then we can compute the expectation in (6) in time $o(2^n)$ (check e.g. [CR20], [CLW20]), which allows us to put L in $\mathsf{NTIME}[o(2^n)]$. This is a contradiction. Hence, we conclude that for infinitely many n, we have $1^n \in L$ but any proof $y \in \{0,1\}^{2^n \mathsf{poly}(n)}$ for 1^n cannot be computed by a circuit of depth $n^{o(1)}$. □

Acknowledgments

We would like to thank William Hoza for asking whether PRGs for $\mathcal{C}$ imply average-case lower bounds against $\mathcal{C}$ under the uniform distribution. We are also grateful to Roei Tell for the observation that part of our equivalence theorem extends to the constant-error case as well (Theorem 14). Lijie Chen is supported by an IBM fellowship. This work received support from the Royal Society University Research Fellowship URF\R1\191059.

References

[AAW10] Eric Allender, Vikraman Arvind, and Fengming Wang. Uniform derandomization from pathetic lower bounds. In *RANDOM-APPROX*, pages 380–393. Springer, 2010.

[ACW16] Josh Alman, Timothy M. Chan, and R. Ryan Williams. Polynomial representations of threshold functions and algorithmic applications. In *FOCS*, pages 467–476, 2016.

[AIK06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. Cryptography in NC^0. *SIAM J. Comput.*, 36(4):845–888, 2006.

[App17] Benny Applebaum. Garbled circuits as randomized encodings of functions: a primer. In *Tutorials on the Foundations of Cryptography*, pages 1–44. Springer International Publishing, 2017.

[AS14] Sergei Artemenko and Ronen Shaltiel. Lower bounds on the query complexity of non-uniform and adaptive reductions showing hardness amplification. *Comput. Complex.*, 23(1):43–83, 2014.

[BHLR19] Abhishek Bhrushundi, Kaave Hosseini, Shachar Lovett, and Sankeerth Rao. Torus polynomials: An algebraic approach to ACC lower bounds. In *ITCS*, pages 13:1–13:16, 2019.

[BT94] Richard Beigel and Jun Tarui. On ACC. *Comput. Complex.*, 4:350–366, 1994.

[BV14] Eli Ben-Sasson and Emanuele Viola. Short PCPs with projection queries. In *ICALP*, pages 163–173, 2014.

[CFL83] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *STOC*, pages 94–99, 1983.

[Che19] Lijie Chen. Non-deterministic quasi-polynomial time is average-case hard for ACC circuits. In *FOCS*, pages 1281–1304, 2019.

[CL21] Lijie Chen and Xin Lyu. Inverse-exponential correlation bounds and extremely rigid matrices from a new derandomized XOR lemma. In *STOC*, 2021.

[CLW20] Lijie Chen, Xin Lyu, and Ryan Williams. Almost-everywhere circuit lower bounds from non-trivial derandomization. In *FOCS*, pages 1–12, 2020.

[CM18] Arkadev Chattopadhyay and Nikhil S. Mande. A short list of equalities induces large sign rank. In *FOCS*, pages 47–58, 2018.

[CR20] Lijie Chen and Hanlin Ren. Strong average-case lower bounds from non-trivial derandomization. In *STOC*, pages 1327–1334, 2020.

[CW19] Lijie Chen and Ryan Williams. Stronger connections between circuit analysis and circuit lower bounds, via PCPs of proximity. In *CCC*, pages 19:1–19:43, 2019.

[For01] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. In *CCC*, pages 100–106, 2001.

[FSUV13] Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. *Theory Comput.*, 9:809–843, 2013.

[GGH+07] Shafi Goldwasser, Dan Gutfreund, Alexander Healy, Tali Kaufman, and Guy N. Rothblum. Verifying and decoding in constant depth. In *STOC*, pages 440–449, 2007.

[GHR92] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov. Majority gates vs. general weighted threshold gates. *Comput. Complex.*, 2:277–300, 1992.

[GKT92] Frederic Green, Johannes Köbler, and Jacobo Torán. The power of the middle bit. In *CCC*, pages 111–117, 1992.

[GNW11] Oded Goldreich, Noam Nisan, and Avi Wigderson. On Yao's XOR-lemma. In *Studies in Complexity and Cryptography*, pages 273–301. 2011.

[GR08] Dan Gutfreund and Guy N. Rothblum. The complexity of local list decoding. In *RANDOM-APPROX*, pages 455–468, 2008.

[GSV18] Aryeh Grinberg, Ronen Shaltiel, and Emanuele Viola. Indistinguishability by adaptive procedures with advice, and lower bounds on hardness amplification proofs. In *FOCS*, pages 956–966, 2018.

[Hås94] Johan Håstad. On the size of weights for threshold gates. *SIAM J. Discret. Math.*, 7(3):484–492, 1994.

[HG91] Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Comput. Complex.*, 1:113–129, 1991.

[HMP+93] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[HV21] Xuangui Huang and Emanuele Viola. Average-case rigidity lower bounds. In *CSR*, 2021.

[IK02] Yuval Ishai and Eyal Kushilevitz. Perfect constant-round secure computation via perfect randomizing polynomials. In *ICALP*, pages 244–256, 2002.

[IM21] Russell Impagliazzo and Sam McGuire. Comparing computational entropies below majority (or: When is the dense model theorem false?). In *ITCS*, pages 2:1–2:20, 2021.

[Imp95] Russell Impagliazzo. Hard-core distributions for somewhat hard problems. In *FOCS*, pages 538–545, 1995.

[IW97] Russell Impagliazzo and Avi Wigderson. $P = BPP$ if E requires exponential circuits: Derandomizing the XOR lemma. In *STOC*, pages 220–229, 1997.

[Kli01] Adam R. Klivans. On the derandomization of constant depth circuits. In *RANDOM-APPROX*, pages 249–260, 2001.

[Kri21] Vaibhav Krishan. Upper bound for torus polynomials. In *CSR*, 2021.

[KW16] Daniel M. Kane and Ryan Williams. Super-linear gate and super-quadratic wire lower bounds for depth-two and depth-three threshold circuits. In *STOC*, pages 633–643, 2016.

[Lev87] Leonid A. Levin. One-way functions and pseudorandom generators. *Comb.*, 7(4):357–363, 1987.

[LTW11] Chi-Jen Lu, Shi-Chun Tsai, and Hsin-Lung Wu. Complexity of hard-core set proofs. *Comput. Complex.*, 20(1):145–171, 2011.

[MW20] Cody D. Murray and R. Ryan Williams. Circuit lower bounds for nondeterministic quasi-polytime from a new easy witness lemma. *SIAM J. Comput.*, 49(5), 2020.

[NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Comput. Complex.*, 4:301–313, 1994.

[NW94] Noam Nisan and Avi Wigderson. Hardness vs randomness. *J. Comput. Syst. Sci.*, 49(2):149–167, 1994.

[Raz87] Alexander A. Razborov. Lower bounds on the size of constant-depth networks over a complete basis with logical addition. *Mathematicheskie Zametki*, 41(4):598–607, 1987.

[RW93] Alexander A. Razborov and Avi Wigderson. $n^{\Omega(\log n)}$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Inf. Process. Lett.*, 45(6):303–307, 1993.

[STV01] Madhu Sudan, Luca Trevisan, and Salil P. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. Syst. Sci.*, 62(2):236–266, 2001.

[SV10] Ronen Shaltiel and Emanuele Viola. Hardness amplification proofs require majority. *SIAM J. Comput.*, 39(7):3122–3154, 2010.

[Tam16] Suguru Tamaki. A satisfiability algorithm for depth two circuits with a sub-quadratic number of symmetric and threshold gates. *Electron. Colloquium Comput. Complex.*, 23:100, 2016.

[Vio05] Emanuele Viola. The complexity of constructing pseudorandom generators from hard functions. *Comput. Complex.*, 13(3-4):147–188, 2005.

[Vio09] Emanuele Viola. Guest column: correlation bounds for polynomials over $\{0,1\}$. *SIGACT News*, 40(1):27–44, 2009.

[Vio19] Emanuele Viola. Constant-error pseudorandomness proofs from hardness require majority. *ACM Trans. Comput. Theory*, 11(4):19:1–19:11, 2019.

[Vio20] Emanuele Viola. New lower bounds for probabilistic degree and AC^0 with parity gates. *Electron. Colloquium Comput. Complex.*, 27:15, 2020.

[Wil13] Ryan Williams. Improving exhaustive search implies superpolynomial lower bounds. *SIAM J. Comput.*, 42(3):1218–1244, 2013.

[Wil18a] Ryan Williams. Limits on representing boolean functions by linear combinations of simple functions: Thresholds, ReLUs, and low-degree polynomials. In *CCC*, pages 6:1–6:24, 2018.

[Wil18b] Ryan Williams. New algorithms and lower bounds for circuits with linear threshold gates. *Theory of Computing*, 14(1):1–25, 2018.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il